A.P. Mishina
I.P. Prokuryakov

Higher Algebra
Linear Algebra
Polynomials
General Algebra

INTERNATIONAL SERIES OF MONOGRAPHS IN

PURE AND APPLIED MATHEMATICS

GENERAL EDITORS: I. N. SNEDDON, M. STARK, K.A.H. GRAVETT AND S. ULAM

VOLUME 80

HIGHER ALGEBRA

Linear Algebra
Polynomials, General Algebra

HIGHER ALGEBRA

Linear Algebra, Polynomials, General Algebra

BY

A. P. MISHINA and I. V. PROSKURYAKOV

Translated from the Russian by
ANN SWINFEN

PERGAMON PRESS
OXFORD · LONDON · EDINBURGH · NEW YORK
PARIS · FRANKFURT

Pergamon Press Ltd., Headington Hill Hall, Oxford
4 & 5 Fitzroy Square, London W.1

Pergamon Press (Scotland) Ltd., 2 & 3. Teviot Place, Edinburgh 1

Pergamon Press Inc., 122 East 55th St., New York 22, N.Y.

Pergamon Press, GmbH, Kaiserstrasse 75, Frankfurt-am-Main

Copyright © 1965
Pergamon Press Ltd.

First English edition 1965

Library of Congress Catalog Card No. 65-21370

This is a translation of the original Russian *Высшая алгебра* (*Vysshaya algebra*), published in 1962 by Fizmatgiz, Moscow

Printed in Poland

CONTENTS

Preface ix
Introduction xi

CHAPTER I

DETERMINANTS AND SYSTEMS OF LINEAR EQUATIONS

1 Determinants 1
1 Permutations and combinations 1
2 The concept of a determinant of any order 4
3 Elementary properties of determinants 6
4 Minors and cofactors of elements. Expansion of a determinant with respect to a row or a column 8
5 Evaluation of a determinant with numerical elements 9
6 Methods of evaluating determinants of the nth order 10
7 The Vandermonde determinant and determinants related to it 13
8 Jacobian determinants. The continuant and the connexion with Fibonacci numbers 15
9 Minors of any order, their algebraic complements and Laplace's theorem 16
10 The expansion of a determinant with respect to a row and a column. Bordered determinants 18
11 The multiplication of determinants. The Binet–Cauchy theorem 21
12 Circulants and the determinants connected with them 23
13 Adjoint (or reciprocal) and associated determinants, their minors and Sylvester's formula 25
14 Symmetric, skew-symmetric and pseudo-symmetric determinants 28
15 Kronecker product of determinants 29
16 Hadamard's inequality and other inequalities connected with determinants 30

2 Systems of Linear Equations 33
1 General information 33
2 Gauss's method of eliminating unknowns 34
3 Cramer's rule 38
4 The rank of a matrix 40
5 The general theory of linear equations 44

6 Systems of homogeneous linear equations. Fundamental systems of solutions 48
7 The connexion between the solutions of homogeneous and non-homogeneous systems 51

Chapter II

Matrices and Quadratic Forms

1 Matrices 52
1 Operations with matrices 52
2 The identity matrix and the inverse matrix 56
3 Powers of matrices. Polynomials in matrices. Commuting matrices 60
4 The connexion between the multiplication of matrices and elementary transformations. Decomposition of a matrix into the product of triangular matrices 63
5 Polynomial matrices, invariant factors and elementary divisors 66
6 Matrix polynomials 73
7 The characteristic roots and characteristic vectors of a matrix 75
8 The minimal polynomial of a matrix 78
9 Similar matrices 81
10 The diagonal form of a matrix 84
11 The Jordan form of a matrix 88
12 The natural form and other canonical forms of a matrix 92
13 Jacobian matrices 97
14 Associated and adjoint matrices 99
15 Kronecker product of matrices 101
16 Orthogonal and unitary matrices 103
17 Symmetric, skew-symmetric and Hermitian matrices 109
18 Normal matrices 115

2 Bilinear and Quadratic Forms 117
1 Equivalence. The diagonal and normal forms. The methods of Lagrange and Jacobi for reduction to diagonal form 117
2 Real quadratic forms. The law of inertia. Indices of inertia and signature 127
3 Positive definite quadratic forms 130
4 Orthogonal transformations of quadratic forms. The reduction of a quadratic form to principal axes 132
5 A pair of forms. A pencil of forms 139
6 Extremal properties of the characteristic roots of a regular pencil of forms 142
7 Complex forms of the second kind. Hermitian forms 145
8 Hänkel forms 149

Chapter III

The Algebra of Polynomials

1 The General Properties of Polynomials 152
1 Definitions. Examples 152
2 Division with a remainder 153
3 Horner's scheme 153
4 The expansion of a polynomial into powers of a difference 155
5 The greatest common divisor of two polynomials 156
6 Mutually prime polynomials 158
7 The representation of $d(x) = (f(x), g(x))$ in the form $d(x) = f(x)u(x) + g(x)v(x)$ 158
8 Eliminating irrationals in the denominator 161
9 The greatest common divisor of several polynomials 161
10 The roots of a polynomial. The separation of multiple roots 162
11 Vieta's formulae 164
12 Common roots of two polynomials. Resultant. Discriminant 164

2 The Interpolation Formulae of Lagrange and Newton 167

3 Location of the Roots of Polynomials. The Decomposition of Polynomials into Factors 171
1 Solution of equations of the second degree 171
2 Solution of equations of the third degree 172
3 Solution of equations of the fourth degree 176
4 Equations of degree higher than the fourth 178
5 Rational roots of polynomials with rational coefficients 178
6 Decomposition of a polynomial into factors of the first and second degree 180
7 Irreducible polynomials 181
8 Decomposition of polynomials into irreducible factors over the field of rational numbers 182
9 Rational functions 186

4 Bounds on the Roots of a Polynomial. Theorems on the Number of Roots 190
1 Bounds on the real roots of a polynomial with real coefficients 190
2 Theorems on the number of real roots 193
3 Bounds on the complex roots of a polynomial 201
4 The number of roots in the left and right half-planes 204
5 Criteria of stability 214
6 The domain of stability 215

5 Polynomials in Several Variables 216
1 Definitions 216
2 Symmetric polynomials 217

3 Expressions for power sums in terms of the elementary symmetric polynomials 220
4 Systems of algebraic equations in several unknowns 221

Chapter IV

General Algebra

1 Groups 224
1 Definition of group. Examples 224
2 Isomorphism of groups 227
3 Homomorphism 227
4 Subgroups. Cyclic groups 228
5 Systems of generators. Increasing sequences of subgroups 230
6 Cosets. Decomposition of a group with respect to a subgroup 231
7 Normal subgroups of a group 233
8 Factor-group 235
9 Groups of permutations 236
10 Direct products of groups 239
11 Abelian groups 241

2 Rings 243
1 Definitions. Examples 243
2 Isomorphism. Homomorphism 245
3 Subrings. Ideals 246
4 Direct sums of rings 247
5 Factor-ring 248

3 Fields. Skew Fields 248
1 Fields 248
2 Skew fields 251

4 Algebras 251

5 Lattices 254
1 Partially ordered sets 254
2 Fundamental definitions 255
3 Distributive and modular lattices 257
4 Boolean algebras 259

Bibliography 261

Index of Special Symbols 263

Subject Index 265

Other Titles in Series 275

PREFACE

In the course of the scientific work of technical specialists, mathematicians, physicists, chemists and others, the situation often arises when it is necessary to obtain information on some mathematical problem without a detailed study of it. It is not always easy to find the relevant books, particularly if the problem is not elementary, and sometimes it is difficult to trace the necessary result in the text of a monograph and to separate it from the preceding account.

It seems to us that the publication of a series of reference books on higher mathematics will be very useful in this respect. The present book is devoted to higher algebra. In it, in Chapters I and II, are collected results relating to linear algebra (matrices, determinants, systems of linear equations, quadratic forms); in Chapter III the algebra of polynomials is presented (operations on polynomials, algebraic equations with one unknown and the properties of their roots, rational functions, polynomials in several variables); Chapter IV is devoted to the most important concepts of algebra as a branch of modern mathematics (groups, rings, fields and skew fields, algebras, lattices).

In the brief Introduction we give some fundamental definitions which are used later in the book, and it is necessary to read it in order to understand the references.

There are no proofs, but the formulation of the results is accompanied by a series of examples to explain their significance. Interested readers may pursue a problem by means of the detailed Contents or the Subject Index at the end of the book.

We must add that the material most often encountered in practical applications is contained in the first three chapters; the fourth chapter will be of most assistance to the reader in supplying an outline of the concepts of modern algebra which may be unfamiliar to him.

In references to a section of the same chapter, the number of the chapter is omitted; similarly, in references to a paragraph of the same section, the number of this section is omitted.

Chapters I and II were written by I. V. Proskuryakov; Chapters III and IV by A. P. Mishina.

P. K. Rashevskii

INTRODUCTION

THROUGHOUT the whole of this book we will encounter the following fundamental concepts: set, number field, matrix. The fundamental concepts and notations of the theory of sets are given in [34], pp. 22–24.

A *number field* K is any set of numbers which contains at least one number distinct from zero, and together with any two numbers a and b contains their sum $a+b$, their difference $a-b$, their product ab and their quotient a/b (if $b \neq 0$). We can verify that the set of all rational numbers forms a number field. In exactly the same way the real numbers and the complex numbers form number fields. These three number fields play a very important part in the whole of mathematics. The field of complex numbers has the following important property: any polynomial of degree $n > 0$ in one variable x with complex (in particular, with real) coefficients has all n of its roots in the same field (see Chapter III). This property of the field of complex numbers is called the property of being algebraically closed. In addition to number fields we will also consider the more general concept of a field whose elements are not necessarily numbers (see Chapter IV). In the first three chapters (unless we specify otherwise) we will only discuss number fields.

A *matrix* is a rectangular array of numbers. To denote a matrix we use curved brackets or double vertical lines, thus:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \qquad \text{or} \qquad \begin{Vmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{Vmatrix}.$$

In the present book we will use the first notation. The numbers which form a matrix are called its elements. The horizontal lines of elements of a matrix are called its rows, and the vertical lines are called its columns. If the number of rows of a matrix is equal to the number of columns, then the matrix is called *square*, and the number of rows is called the *order* of this matrix. In what follows, for brevity, by matrices of a given order we shall mean square matrices of this order. In the general form of notation the elements of a matrix are denoted

usually by one letter with two indices, of which the first denotes the number of the row, and the second the number of the column, for instance, a matrix of order n in the general form is written thus:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & & & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

(this is read: "a one, one, a one, two", and so on, and not "a eleven, a twelve"). The series of elements of a square matrix lying in the line joining the upper left-hand corner to the lower right-hand corner is called the *principal* (or first) diagonal, and that lying in the line joining the upper right-hand corner to the lower left-hand corner is called the *indirect* (or second) diagonal of the matrix.

CHAPTER I

DETERMINANTS AND SYSTEMS OF LINEAR EQUATIONS

1. Determinants

1. Permutations and combinations.

A *combination* of the n numbers $1, 2, \ldots, n$ (or any n distinct symbols $a_1, a_2, \ldots, a_n$) is any arrangement of these numbers (or symbols) in a definite order. Because the given n symbols can be indexed by means of the numbers $1, 2, \ldots, n$, the study of the combinations of any n symbols reduces to the study of the combinations of these numbers. The total number of combinations of n numbers is equal to $1 \times 2 \times 3 \ldots n = n!$ (read: "n factorial").

EXAMPLE 1. The combinations of the numbers 1, 2, 3 are: 123, 132, 213, 231, 312, 321. The number of these is $3! = 6$.

We say that two numbers in a combination occur *inversely* if the larger number stands before the smaller; if, however, the smaller number stands before the larger, we say that they occur *in order*.

Method of calculating the number of inversions. We read the numbers in the combination in the order in which they are written (from left to right), for each of the numbers we count how many numbers less than the given one stand to the right of it, and add together all the numbers obtained.

EXAMPLE 2. In the combination 528371964 the number of inversions is equal to $4 + 1 + 5 + 1 + 3 + 2 + 1 = 17$.

A combination is called *even* or *odd* according to whether the number of inversions in it is even or odd.

A transposition is the operation of interchanging two numbers in a combination while leaving the other numbers fixed. The transposition of the numbers in i and j is denoted by (i, j).

We can pass from any combination of n numbers to any other combination of the same numbers by means of a series of transpositions, and we need not more than $n-1$ transpositions.

EXAMPLE 3. We can pass from the combination 25134 to the combination 42513 by means of the four transpositions: (2, 4), (2, 5), (1, 5), (1, 3).

All the $n!$ combinations of the numbers $1, 2, \ldots, n$ can be set out one after the other (without omissions and without repetitions) in such a way that each successive combination is obtained from the preceding one by means of one transposition. Every transposition changes the parity of the combination. For any $n \geqslant 2$ the number of even combinations of n numbers is equal to the number of odd combinations, i.e. $1/2\ n!$

A *permutation* of the n numbers $1, 2, \ldots, n$ or a *permutation of degree n* is a one–one mapping of the totality of these numbers on itself, i.e. a mapping under which each of the numbers from 1 to n corresponds to a certain one of these numbers and two distinct numbers always correspond to two distinct numbers. A permutation is written as two rows between brackets, where to every number in the upper row there corresponds the number standing below it. For instance $\begin{pmatrix} 2 & 1 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}$ denotes the permutation in which: $1 \to 1, 2 \to 3, 3 \to 4, 4 \to 2$.

Depending on the arrangement of the numbers in the upper row, the same permutation can be written in many ways. For instance, the expressions

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 3 & 2 & 1 \\ 1 & 3 & 2 \end{pmatrix}$$

denote the same permutation, under which 1 goes into 2, 2 into 3, 3 into 1. Every permutation of n numbers can be written in $n!$ different ways. The number of distinct permutations of n elements is also equal to $n!$

A permutation is called *even* if the total number of inversions in both its rows is even, and *odd* if it is odd. In other words, a permutation is even if its rows have the same parity, and odd if they have opposite parity. The parity of a permutation does not depend on the way in which it is written in the two rows, i.e. all the expressions for the same permutation have the same parity.

EXAMPLE 4. The permutation

$$\begin{pmatrix} 2 & 3 & 4 & 1 \\ 3 & 1 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}$$

contains six inversions in the first expression and two in the second, and so is even.

The number of even permutations of n elements is equal to the number of odd ones, i.e. $1/2\, n!$ (if $n \geqslant 2$).

There is another method of defining the parity of a permutation. A *cycle* is a sequence of several numbers in which the first number in the given permutation goes into the second, the second into the third, and so on, and the last into the first. A cycle is denoted by enclosing its numbers in brackets. If a number is carried into itself, then it generates a cycle by itself. Cycles which have no numbers in common are called *independent*. Any permutation can be decomposed into independent cycles.

EXAMPLE 5. $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 5 & 1 & 4 & 2 & 3 \end{pmatrix} = (163)\quad (25)\quad (4).$

The *decrement* of a permutation is the difference between the number n of the elements permuted and the number k of the cycles in its decomposition: $d = n - k$. The parity of a permutation coincides with the parity of its decrement. Thus, in Example 5, $n = 6$, $k = 3$, $d = 3$, i.e. the permutation is odd.

The smallest number of transpositions necessary for the transition from one combination of the n numbers $a_1, a_2, \ldots, a_n$ to another combination of the same numbers $b_1, b_2, \ldots, b_n$ is equal to the decrement of the permutation

$$\begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ b_1 & b_2 & \ldots & b_n \end{pmatrix}$$

(see [20], Problem 183). Hence, for any combination of n numbers there exist combinations of the same numbers which it is impossible to transform to the given combination by means of less than $n - 1$ transpositions.

We can introduce the operation of multiplying permutations by giving another definition of a cycle as a permutation, and interpreting the decomposition of a permutation into independent cycles as a representation of it in the form of the product of these cycles. For this see Chapter IV, section 1, subsections 1 and 9, and also [11], pp. 33–36.

2. The concept of a determinant of any order.

Suppose that we are given a square matrix of order n:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} = (a_{ij})_1^n.$$

The number obtained from the formulae is called a determinant of order n or the determinant of the matrix **A**.

$$|\mathbf{A}| = |a_{ij}|_1^n = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} = \sum (-1)^{s+t} a_{i_1 j_1} a_{i_2 j_2} \dots a_{i_n j_n}$$

$$= \sum (-1)^k a_{1k_1} a_{2k_2} a_{3q} \dots a_{nk_n} = \sum (-1)^l a_{l_1 1} a_{l_1 1} \dots a_{l_1 1}.$$

Here the first three expressions are notations for the determinant; the first sum is taken over all the distinct permutations

$$\begin{pmatrix} i_1 & i_2 & \dots & i_n \\ j_1 & j_2 & \dots & j_n \end{pmatrix}, \qquad (*)$$

where s is the number of inversions in the upper, and t the number of inversions in the lower row of this permutation, the second sum is taken over all the combinations $k_1, k_2, \dots, k_n$ and k is the number of inversions in this combination. The third sum has a similar meaning, and all three sums are identically equal. Their summands are called the *terms of the determinant*; every term of the determinant is equal to the product of n elements of the matrix, taken one from each row and one from each column, where this product is taken with the positive sign if the permutation (*) is even, and with the opposite sign if it is odd. A determinant of the first order is equal to its own single element.

The number of terms of a determinant of order n is equal to $n!$ The elements, rows, columns and so on of the matrix **A** are called, the elements, rows, columns and so on of its determinant $|\mathbf{A}|$.

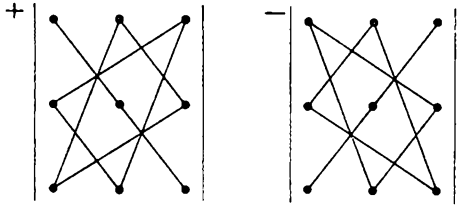
EXAMPLE 1. A determinant of the second order:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11} a_{22} - a_{12} a_{21}$$

EXAMPLE 2. A determinant of the third order:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{21}a_{32}a_{13} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{23}a_{32}a_{11}.$$

This expression is obtained by the triangle rule (Sarrius's rule), and may be clarified by means of the following schemes, in which the elements occurring in one product with the given sign are joined by straight lines:



EXAMPLE 3. The product of the elements of the principal diagonal $a_{11}\,a_{22}\,\ldots a_{nn}$ for any n has the plus sign in a determinant of order n, because neither the first nor any other indices have inversions.

EXAMPLE 4. To determine the sign of the product of the elements in the indirect diagonal $a_{1n}a_{2,n-1}a_{3,n-2}\ldots a_{n1}$ of a determinant of order n.

Because the indices of the rows go in increasing order, the sign is determined by the parity of the combination of the indices of the columns n, $n-1, n-2, \ldots, 1$, in which any two numbers form an inversion, so that the number of all the inversions is equal to $C_n^2 = n(n-1)/2$, and the product of the elements of the indirect diagonal in a determinant of order n has the sign $(-1)^{n(n-1)/2}$. When $n = 4k, 4k+1$ we obtain the plus sign, and when $n = 4k+2, 4k+3$ we obtain the minus sign. Thus, for the orders 2, 3 we obtain the minus sign, for 4, 5 the plus sign, for 6, 7 the minus sign, for 8, 9 the plus sign, and so on.

EXAMPLE 5. To evaluate the determinant

$$d = \begin{vmatrix} 0 & a_{12} & a_{13} & 0 \\ a_{21} & 0 & 0 & a_{24} \\ a_{31} & 0 & 0 & a_{34} \\ 0 & a_{42} & a_{43} & 0 \end{vmatrix}.$$

If we do not take zero summands, then from the first row we need only take a_{12} or a_{13}. If we take a_{12}, then from the second row we need only take a_{21} or a_{24}; if we take a_{21}, then from the third row we need to take a_{34} and from the fourth a_{43}. Picking out all such possibilities and taking the signs into account, we obtain:

$$d = a_{12}a_{21}a_{34}a_{43} - a_{12}a_{24}a_{31}a_{43} - a_{13}a_{21}a_{34}a_{42} + a_{13}a_{24}a_{31}a_{42}.$$

To calculate more complicated determinants by means of the immediate application of the definition would not

always be convenient. For this we make use of special methods (see subsections 5 and 6).

3. Elementary properties of determinants.

Transposition of a matrix (and also of a determinant) is the interchange of the rows and columns while preserving their numbering. Thus the rows of the given matrix will be the columns of the transposed matrix in the same order and vice versa.

EXAMPLE 1. By transposing the matrix $\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ we obtain the matrix

$$\mathbf{A}' = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

In the case of a square matrix (or in the case of a determinant), transposition reduces to rotating the matrix (or the determinant) through 180° around the principal diagonal.

The *sum of several rows* of the same length is the row each of whose elements is equal to the sum of the corresponding elements of the given rows. The *product of a row by a number* is the row each of whose elements is obtained from the corresponding element of the given row by multiplying it by the given number. A *linear combination* of several rows of the same length is a row which is equal to the sum of products of the given rows by certain numbers, called the coefficients of this linear combination. If one row is a linear combination of others, then we also say that it *can be expressed linearly* in terms of these rows.

Several rows of the same length are called *linearly independent* if none of them can be expressed linearly in terms of the rest.

EXAMPLE 2. The equation

$$(1, -1, -3, -5) = 3(1, 1, 1, 1) - 2(1, 2, 3, 4)$$

means that the first row is a linear combination of the other two.

Elementary properties of determinants.

(1) Transposition of a determinant does not alter its value.

(2) If all the elements of a particular row (or column) of a determinant are zero, then the determinant is zero.

(3) Interchanging any two rows (or columns) of a determinant changes its sign.

(4) If two rows (or two columns) of a determinant are identical, then it is zero.

(5) If all the elements of one row (or column) of a determinant are multiplied by the same number, then the whole determinant is multiplied by this number. In other words, a common factor of the elements of a row (or column) can be taken outside the determinant sign.

(6) If the elements of one row of a determinant are proportional to the corresponding elements of another row, then the determinant is equal to zero. This is true for columns.

(7) If each of the elements of the ith row of a determinant can be represented in the form of the sum of k summands, then the whole determinant can also be represented in the form of the sum of k determinants, in which all the rows, except the ith, are the same as those in the original determinant, and the ith row in the first determinant consists of the first summands, in the second determinant of the second, and so on. The same is true for columns.

For instance,

$$\begin{vmatrix} 5 & -3 & 1 \\ 3 & 2 & -7 \\ 4 & 0 & -2 \end{vmatrix} = \begin{vmatrix} 5 & -3 & 1 \\ 1 & 1 & -6 \\ 4 & 0 & -2 \end{vmatrix} + \begin{vmatrix} 5 & -3 & 1 \\ 2 & 1 & -1 \\ 4 & 0 & -2 \end{vmatrix}.$$

(8) If we add one or several other rows to a given row of a determinant, and they are multiplied by any number, then the value of the determinant is unaltered. The same is true for columns. In particular, we can add to one row (column) (or take away from it) another row (column).

(9) If at least one row of a determinant can be expressed linearly in terms of other rows of it, then the determinant is zero. Conversely, if a determinant of order $n \geqslant 2$ is zero, then at least one of its rows can be expressed linearly in terms of the other rows. The same is true for columns.

EXAMPLE 3. A determinant D, in which any two elements a_{jk} and a_{kj} which lie in symmetric positions with respect to the principal diagonal are complex conjugates, is real.

In fact, let $D = a + bi$. By transposing D, we obtain a determinant D' all of whose elements are conjugate to the corresponding elements of the determinant D. But the determinant is equal to the sum of products (with the proper signs) of its elements, and by the properties of complex numbers conjugate to the sum and product of several numbers, we obtain: $D' = a - bi$. Because $D = D'$, then $a + bi = a - bi$, whence $b = 0$ and $D = a$, a real number.

EXAMPLE 4. A determinant D is called *skew-symmetric* if any two elements which lie in symmetric positions with respect to the principal diagonal differ only in sign, i.e. $a_{jk} = -a_{kj}$ (in particular, the elements of the principal diagonal are zero). A skew-symmetric determinant D of odd order n is zero. In fact, by removing a factor -1 from every row of the determinant D we obtain the transposed determinant, equal to D, i.e. $D = (-1)^n D$ and $D = 0$ because n is odd.

4. Minors and cofactors of elements. Expansion of a determinant with respect to a row or a column.

The *minor* M_{ij} of the element a_{ij} of the determinant D of order $n \geqslant 2$ is the determinant of order $n-1$ obtained from D by crossing out the ith row and the jth column at whose intersection stands the given element a_{ij}.

The *cofactor* of the element a_{ij} is the number $A_{ij} = (-1)^{i+j} M_{ij}$.

A determinant is equal to the sum of the products of the elements of any one of its rows (or columns) by their cofactors. In particular, if all the elements of a row (or column), except one, are zero, then the determinant is equal to this element multiplied by its cofactor.

EXAMPLE 1. The minor of the element a_{23} of the determinant

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}$$

is

$$M_{23} = \begin{vmatrix} a_{11} & a_{12} & a_{14} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{vmatrix}.$$

The cofactor of the element a_{23} is $A_{23} = -M_{23}$.

EXAMPLE 2. To evaluate the determinant

$$D = \begin{vmatrix} 2 & -3 & 4 & 1 \\ 4 & -2 & 3 & 2 \\ a & b & c & d \\ 3 & -1 & 4 & 3 \end{vmatrix}.$$

Expanding with respect to the third row, we find

$$D = a(-1)^4 \begin{vmatrix} -3 & 4 & 1 \\ -2 & 3 & 2 \\ -1 & 4 & 3 \end{vmatrix} + b(-1)^5 \begin{vmatrix} 2 & 4 & 1 \\ 4 & 3 & 2 \\ 3 & 4 & 3 \end{vmatrix} + c(-1)^6 \begin{vmatrix} 2 & -3 & 1 \\ 4 & -2 & 2 \\ 3 & -1 & 3 \end{vmatrix}$$

$$+ d(-1)^7 \begin{vmatrix} 2 & -3 & 4 \\ 4 & -2 & 3 \\ 3 & -1 & 4 \end{vmatrix} = 8a + 15b + 12c - 19d.$$

5. Evaluation of a determinant with numerical elements.

The evaluation of a determinant of order n can be reduced to the evaluation of one determinant of order $n-1$. Suppose that we have to evaluate the determinant

$$D = \begin{vmatrix} a_{11} & a_{12} \dots a_{1n} \\ a_{21} & a_{22} \dots a_{2n} \\ \cdot & \cdot \quad \cdot \quad \cdot \\ a_{n1} & a_{n2} \dots a_{nn} \end{vmatrix}.$$

If all the elements of the first column are equal to zero, then $D = 0$ (see Property 2, p. 6); if $a_{11} = 0$, but $a_{k1} \neq 0$, then, interchanging the first and kth rows, we obtain in the upper left-hand corner an element distinct from zero. Thus, without loss of generality, we can assume that $a_{11} \neq 0$. To the second row we add the first multiplied by $-a_{21}/a_{11}$, to the third row we add the first multiplied by $-a_{31}/a_{11}$, ..., to the nth row we add the first multiplied by $-a_{n1}/a_{11}$. The determinant does not change its value (see p. 7, Property 8), and assumes the form

$$D = \begin{vmatrix} a_{11} & a_{12} \dots a_{1n} \\ 0 & b_{22} \dots b_{2n} \\ \cdot & \cdot \quad \cdot \quad \cdot \\ 0 & b_{n2} \dots b_{nn} \end{vmatrix} = a_{11} |b_{ij}|_1^{n-1} \text{ (see subsection 4).}$$

EXAMPLE. To evaluate the determinant

$$D = \begin{vmatrix} 6 & -5 & 8 & 4 \\ 9 & 7 & 5 & 2 \\ 7 & 5 & 3 & 7 \\ -4 & 8 & -8 & -3 \end{vmatrix}.$$

We have

$$D = \begin{vmatrix} -12 & -19 & -2 & 0 \\ 9 & 7 & 5 & 2 \\ -20 & -16 & -12 & 1 \\ 5 & 15 & -3 & -1 \end{vmatrix} = \begin{vmatrix} -12 & -19 & -2 & 0 \\ 19 & 37 & 1 & 0 \\ -15 & -1 & -15 & 0 \\ 5 & 15 & -3 & -1 \end{vmatrix}$$

$$= \begin{vmatrix} -12 & -19 & -2 \\ 19 & 37 & -1 \\ -15 & -1 & -15 \end{vmatrix} = - \begin{vmatrix} -10 & -19 & 283 \\ 20 & 37 & -556 \\ 0 & -1 & 0 \end{vmatrix}$$

$$= - \begin{vmatrix} -10 & 283 \\ 20 & -556 \end{vmatrix} = - \begin{vmatrix} -10 & 283 \\ 0 & 10 \end{vmatrix} = 100.$$

6. Methods of evaluating determinants of the nth order.

The method of evaluating a determinant of given order with numerical elements presented in the preceding section is not convenient in the case of determinants whose elements are letters, or for determinants with numerical elements but of large order n. There is no general method for evaluating such determinants if we do not consider the evaluation of a determinant of given order immediately by its definition (p. 4). Various methods of evaluation are applicable to determinants of one special type or another, which can be given more simply than the expression obtained directly from the definition. Below we give some of these methods.

(1) *The method of reduction to triangular form.* By using the properties of a determinant (see pp. 6–7), we can reduce it to the so-called "triangular" form, when all the elements which stand on one side of the principal diagonal are zero. The determinant of triangular form thus obtained is equal to the product of the elements of the principal diagonal, i.e.

$$D = \begin{vmatrix} a_{11} & a_{12} & a_{13} \dots & a_{1n} \\ 0 & a_{22} & a_{23} \dots & a_{2n} \\ 0 & 0 & a_{33} \dots & a_{3n} \\ \cdot & \cdot & \cdot \cdot \cdot & \cdot \\ 0 & 0 & 0 \dots & a_{nn} \end{vmatrix} = a_{11} a_{22} a_{33} \dots a_{nn} .$$

The case when the zeros stand on the right of the principal diagonal can be reduced to the preceding case by transposition, or can be considered analogously. If it is more convenient to obtain the zeros on one side of the indirect diagonal, then, by reversing the order of the rows or columns of such a determinant D, we find that $D = (-1)^{n(n-1)/2} D_1$, where n is the order of D and D_1 is in triangular form.

EXAMPLE 1.

$$D = \begin{vmatrix} n & n-1 \dots 4 & 3 & 2 & 1 \\ n & n-1 \dots 4 & 3 & 3 & 1 \\ n & n-1 \dots 4 & 5 & 2 & 1 \\ n & n-1 \dots 7 & 3 & 2 & 1 \\ \cdot & \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot & \cdot & \cdot & \cdot \\ 2n-1 & n-1 \dots 4 & 3 & 2 & 1 \end{vmatrix}$$

(in the indirect diagonal stand the first n odd numbers). Subtracting the first row from all the remaining ones, we find

$$D = \begin{vmatrix} n & n-1 & \dots & 4 & 3 & 2 & 1 \\ 0 & 0 & \dots & 0 & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 2 & 0 & 0 \\ 0 & 0 & \dots & 3 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ n-1 & 0 & \dots & 0 & 0 & 0 & 0 \end{vmatrix}$$

$$= (-1)^{n(n-1)/2} \begin{vmatrix} 1 & 2 & 3 & 4 & \dots & n-1 & n \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 2 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 3 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \dots & 0 & n-1 \end{vmatrix} = (1)^{n(-1)/2}(n-1)!$$

EXAMPLE 2. To evaluate the determinant of order n:

$$D = \begin{vmatrix} a & b & b & \dots & b \\ b & a & b & \dots & b \\ b & b & a & \dots & b \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ b & b & b & \dots & a \end{vmatrix}.$$

To the first column we add all the remaining ones:

$$D = \begin{vmatrix} a+(n-1)b & b & b & \dots & b \\ a+(n-1)b & a & b & \dots & b \\ a+(n-1)b & b & a & \dots & b \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a+(n-1)b & b & b & \dots & a \end{vmatrix}$$

(subtract the first row from all the remaining ones)

$$= \begin{vmatrix} a+(n-1)b & b & b & \dots & b \\ 0 & a-b & 0 & \dots & 0 \\ 0 & 0 & a-b & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \dots & a-b \end{vmatrix} = [a+(n-1)b](a-b)^{n-1}.$$

(2) *The method of recurrence relations.* A given determinant D_n of order n, after transformations and expansions with respect to a row or column, can be expressed in terms of determinants of the same form, but of lower orders. The equation obtained is called a *recurrence relation.* We evaluate as many determinants of the given form of the lowest orders as occur on the right-hand side of the recurrence relation. In addition we evaluate the determinants of the following orders, by using the recurrence relation, until we succeed in determining a general law for the expressions obtained. Then, by using the recurrence relation, we can prove this law for D_n by induction with respect to n.

EXAMPLE 3. To evaluate the determinant

$$J_n = \begin{vmatrix} 3 & 2 & 0 & 0 & \dots & 0 & 0 \\ 1 & 3 & 2 & 0 & \dots & 0 & 0 \\ 0 & 1 & 3 & 2 & \dots & 0 & 0 \\ 0 & 0 & 1 & 3 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \dots & 1 & 3 \end{vmatrix}.$$

The recurrence relation is:

$$J_n = 3J_{n-1} - 2J_{n-2}\,.$$

We find that $J_1 = 3$, $J_2 = 7$, $J_3 = 3J_2 - 2J_1 = 15$, $J_4 = 3J_3 - 2J_2 = 31$. The determinants which we have found can be evaluated by means of the formula $J_m = 2^{m+1} - 1$, and we will prove its validity by induction with respect to the number m. For $m = 1$ the formula is true; presupposing that it is true for all $m < n$, in particular, setting $J_{n-1} = 2^n - 1$, $J_{n-2} = 2^{n-1} - 1$, we will have from the recurrence relation

$$J_n = 3J_{n-1} - 2J_{n-2} = 3\,(2^n - 1) - 2\,(2^{n-1} - 1) = 2^{n-1} - 1\,,$$

which is what we were required to prove.

(3) *A special case of recurrence relations.* Suppose that the recurrence relation has the form

$$D_n = pD_{n-1} + qD_{n-2}\,,$$

where p and q are constant numbers (i.e. they do not depend on n). In this case we can deduce a formula for evaluating D_n ([20], pp. 33, 34).

If $q = 0$, then $D_n = p^{n-1} D_1$, where D_1 is the determinant of the first order of the given form. If $q \neq 0$, then we solve the quadratic equation $x^2 - px - q = 0$. Let α and β be its roots. If $\alpha \neq \beta$, then $D_n = c_1 \alpha^n + c_2 \beta^n$, where

$$c_1 = \frac{D_2 - \beta D_1}{\alpha\,(\alpha - \beta)}\,, \qquad c_2 = -\frac{D_2 - \alpha D_1}{\beta\,(\alpha - \beta)}\,.$$

D_1, D_2 are the determinants of the first and second orders of the given form. Expressions for c_1 and c_2 can be found immediately from the equations

$$D_1 = c_1 \alpha + c_2 \beta\,, \quad D_2 = c_1 \alpha^2 + c_2 \beta^2.$$

If $q \neq 0$, but $\alpha = \beta$, then

$$D_n = \alpha^n\,[(n-1)\,c_1 + c_2]\,,$$

where

$$c_1 = \frac{D_2 - \alpha D_1}{\alpha^2} \quad \text{and} \quad c_2 = \frac{D_1}{\alpha}\,.$$

EXAMPLE 4. To evaluate the Jacobian determinant (p. 15) of order n:

$$J_n = \begin{vmatrix} 5 & 3 & 0 & 0 & \dots & 0 & 0 \\ 2 & 5 & 3 & 0 & \dots & 0 & 0 \\ 0 & 2 & 5 & 3 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \dots & 2 & 5 \end{vmatrix}.$$

The recurrence relation is

$$J_n = 5J_{n-1} - 6J_{n-2}.$$

We solve the equation $x^2 - 5x + 6 = 0$; its roots are $\alpha = 2$, $\beta = 3$. In addition we find that $J_1 = 5$, $J_2 = 19$, so that for c_1 and c_2 we have the system of equations

$$2c_1 + 3c_2 = 5, \quad 4c_1 + 9c_2 = 19,$$

whence $c_1 = -2$, $c_2 = 3$ and this means that

$$J_n = c_1\alpha^n + c_2\beta^n = 3^{n+1} - 2^{n+1}.$$

7. The Vandermonde determinant and determinants related to it.

(1) A *Vandermonde determinant* is a determinant of the form

$$D_n = \begin{vmatrix} 1 & x_1 & x_1^2 & x_1^3 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & x_2^3 & \dots & x_2^{n-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_n & x_n^2 & x_n^3 & \dots & x_n^{n-1} \end{vmatrix}.$$

It can be evaluated by means of the formula

$$D_n = (x_2 - x_1)(x_3 - x_1)\dots(x_n - x_1)(x_3 - x_2)(x_4 - x_2)\dots$$
$$\dots(x_n - x_2)\dots(x_n - x_{n-1}) = \prod_{n \geqslant i > k \geqslant 1} (x_i - x_k)$$

(see [11], p. 50, or [20], p. 32).

EXAMPLE 1.

$$D = \begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 3 & \dots & n+1 \\ 1 & 2^2 & 3^2 & \dots & (n+1)^2 \\ 1 & 2^3 & 3^3 & \dots & (n+1)^3 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 2^n & 3^n & \dots & (n+1)^n \end{vmatrix}.$$

This determinant is obtained by transposition from the Vandermonde determinant with defining elements

$$x_1 = 1, \quad x_2 = 2, \dots, x_{n+1} = n+1.$$

Therefore

$$D = (2-1)(3-1)\dots[(n+1)-1]$$
$$\times (3-2)(4-2)\dots[(n+1)-2]\dots[(n+1)-n] = 1!\,2!\,3!\dots n!$$

EXAMPLE 2.

$$D = \begin{vmatrix} a_1^n & a_1^{n-1}b_1 & a_1^{n-2}b_1^2 & \dots & b_1^n \\ a_2^n & a_2^{n-1}b_2 & a_2^{n-2}b_2^2 & \dots & b_2^n \\ \dots & \dots & \dots & \dots & \dots \\ a_{n+1}^n & a_{n+1}^{n-1}b_{n+1} & a_{n+1}^{n-2}b_{n+1}^2 & \dots & b_{n+1}^n \end{vmatrix}.$$

Taking outside the determinant a factor a_1^n from the first row, a factor a_2^n from the second row, and so on, we arrive at the Vandermonde determinant with the defining elements $b_1/a_1, b_2/a_2, \dots, b_{n+1}/a_{n+1}$. Therefore

$$D = a_1^n a_2^n \dots a_{n+1}^n \prod_{n+1 \geqslant i > k \geqslant 1} \left(\frac{b_i}{a_i} - \frac{b_k}{a_k}\right) = \prod_{n+1 \geqslant i > k \geqslant 1} (a_k b_i - a_i b_k).$$

(2) A determinant of the form

$$D = \begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{p-1} & x_1^{p+1} & \dots & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^{p-1} & x_2^{p+1} & \dots & x_2^n \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^{p-1} & x_n^{p+1} & \dots & x_n^n \end{vmatrix} \quad (1 \leqslant p \leqslant n-1)$$

can be reduced to a Vandermonde determinant. If we denote by s_{n-p} the sum of all possible products of the elements $x_1, x_2, \dots, x_n$, taken $n-p$ at a time, and by D_n the Vandermonde determinant of these elements, then $D = s_{n-p} D_n$ ([20], Problems 344, 346).

EXAMPLE 3.

$$D = \begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-2} & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-2} & x_2^n \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-2} & x_n^n \end{vmatrix}$$
$$= (x_1 + x_2 + \dots + x_n) \prod_{n \geqslant i > k \geqslant 1} (x_i - x_k).$$

Here $p = n-1$.

(3) A *generalized Vandermonde determinant* is a determinant

$$D_n = \begin{vmatrix} a_1^{\alpha_1} & a_1^{\alpha_2} & \dots & a_1^{\alpha_n} \\ a_2^{\alpha_1} & a_2^{\alpha_2} & \dots & a_2^{\alpha_n} \\ \dots & \dots & \dots & \dots \\ a_n^{\alpha_1} & a_n^{\alpha_2} & \dots & a_n^{\alpha_n} \end{vmatrix}$$

where $a_1, a_2, \ldots, a_n$ are positive and $\alpha_1, \alpha_2, \ldots, \alpha_n$ are real numbers. This determinant is equal to zero if and only if either amongst the numbers a_i or amongst the numbers α_i some are equal. It is sufficient to consider the case when $0 < a_1 < a_2 < \ldots < a_n$ and $\alpha_1 < \alpha_2 < \ldots < \alpha_n$.

For this case we can prove that $D_n > 0$ (see [7], p. 93).

8. Jacobian determinants. The continuant and the connexion with Fibonacci numbers.

For Jacobian matrices see Chapter II, section 1, subsection 13. A *Jacobian determinant* is a determinant of the form

$$J_n = \begin{vmatrix} a_1 & b_1 & 0 & 0 & \ldots & 0 & 0 \\ -c_2 & a_2 & b_2 & 0 & \ldots & 0 & 0 \\ 0 & -c_3 & a_3 & b_3 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots & a_{n-1} & b_{n-1} \\ 0 & 0 & 0 & 0 & \ldots & -c_n & a_n \end{vmatrix}.$$

Expanding this with respect to the last column, we find the recurrence relation $J_n = a_n J_{n-1} + b_{n-1} c_n J_{n-2}$. We can prove the following rule for forming the expanded expression for J_n. Take the sum of all possible products one of which is $a_1\, a_2\, a_3 \ldots a_n$, and the others are obtained from it by replacing one or several pairs of neighbouring factors $a_i\, a_{i+1}$ by $b_i\, c_{i+1}$.

EXAMPLE 1.

$$J_5 = a_1a_2a_3a_4a_5 + b_1c_2a_3a_4a_5 + a_1b_2c_3a_4a_5 + a_1a_2b_3c_4a_5 + a_1a_2a_3b_4c_5 + b_1c_2b_3c_4a_5 + b_1c_2a_3b_4c_5 + a_1b_2c_3b_4c_5.$$

A special case of a Jacobian determinant is the so-called continuant

$$(a_1a_2a_3 \ldots a_n) = \begin{vmatrix} a_1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ -1 & a_2 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & a_3 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots & a_{n-1} & 1 \\ 0 & 0 & 0 & 0 & \ldots & -1 & a_n \end{vmatrix}.$$

Its name is explained by the connexion with continued fractions, expressed by the equation

$$\frac{(a_1 a_2 a_3 \dots a_n)}{(a_2 a_3 \dots a_n)} = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots + \cfrac{1}{a_n}}}$$

(see [30], p. 29, or [20], Problem 539). It is equal to the sum of the product $a_1 a_2 \dots a_n$ and all possible products obtained from it by crossing out pairs of neighbouring factors (for even n the result of crossing out all n cofactors is regarded as being equal to 1).

EXAMPLE 2.

$$(a_1 a_2 a_3 a_4) = a_1 a_2 a_3 a_4 + a_3 a_4 + a_1 a_4 + a_1 a_2 + 1\,;$$

$$(a_1 a_2 a_3 a_4 a_5) = a_1 a_2 a_3 a_4 a_5 + a_3 a_4 a_5 + a_1 a_4 a_5 + a_1 a_2 a_5 + a_1 a_2 a_3 + a_5 + a_3 + a_1\,.$$

The special form of the continuant

$$u_n = \begin{vmatrix} 1 & 1 & 0 & 0 \dots & 0 & 0 \\ -1 & 1 & 1 & 0 \dots & 0 & 0 \\ 0 & -1 & 1 & 1 \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 \dots & -1 & 1 \end{vmatrix}$$

is equal to the nth term in the so-called *Fibonacci series*

$$1, 2, 3, 5, 8, 13, 21, \dots,$$

in which every term, beginning with the third, is equal to the sum of the two preceding terms ([20], Problem 365).

For other determinants of special forms see section 1, subsections 7, 12, 13, 14, and also [20] and [24].

9. Minors of any order, their algebraic complements and Laplace's theorem.

A *minor of the kth order* in the determinant D of order n $(1 \leqslant k \leqslant n)$ is a determinant M consisting of the elements standing at the intersection of any k rows and any k columns of the determinant D. In particular, the minor of the nth order in D is equal to D itself. A minor of zero order is by definition equal to one. The *complementary minor* for the

minor M in the determinant D is the minor M' obtained from D by crossing out the k rows and k columns in which the given minor M stands. The *algebraic complement* of the minor M in the determinant D is its complementary minor M' taken with the plus sign if the sum of the numbers of the rows and columns of the determinant D in which the original minor M stands is even, and with the minus sign if it is odd.

EXAMPLE 1. In the determinant $D_4 = |a_{ij}|_1^4$ for the minor of the second order

$$M = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$$

the complementary minor will be

$$M' = \begin{vmatrix} a_{12} & a_{14} \\ a_{42} & a_{44} \end{vmatrix}$$

and the algebraic complement is $A = (-1)^9 M' = -M'$.

Laplace's theorem. Suppose that any k rows (or columns) are singled out in the determinant D of order n $(1 \leqslant k \leqslant n-1)$. Then the determinant D is equal to the sum of the products of all the minors of the kth order in the k rows (or columns) singled out, multiplied by their algebraic complements (see [11], p. 51).

For the generalized theorem of Laplace in the case of an expansion with respect to several groups of rows (or columns) see [20], Problem 466.

EXAMPLE 2. To evaluate the determinant

$$D = \begin{vmatrix} 2 & 1 & 4 & 3 & 5 \\ 3 & 4 & 0 & 5 & 0 \\ 3 & 4 & 5 & 2 & 1 \\ 1 & 5 & 2 & 4 & 3 \\ 4 & 6 & 0 & 7 & 0 \end{vmatrix}.$$

Here there are ten minors of the order 2 standing in the second and fifth rows, but only three of them are distinct from zero. Expanding with respect to these two rows, we find

$$D = (-1)^{10} \begin{vmatrix} 3 & 4 \\ 4 & 6 \end{vmatrix} \cdot \begin{vmatrix} 4 & 3 & 5 \\ 5 & 2 & 1 \\ 2 & 4 & 3 \end{vmatrix} + (-1)^{12} \begin{vmatrix} 3 & 5 \\ 4 & 7 \end{vmatrix} \cdot \begin{vmatrix} 1 & 4 & 5 \\ 4 & 5 & 1 \\ 5 & 2 & 3 \end{vmatrix}$$

$$+ (-1)^{13} \begin{vmatrix} 4 & 5 \\ 6 & 7 \end{vmatrix} \cdot \begin{vmatrix} 2 & 4 & 5 \\ 3 & 5 & 1 \\ 1 & 2 & 3 \end{vmatrix} = 2 \cdot 49 + 1 \cdot (-100) + 2 \cdot (-1) = -4.$$

In particular, if in the determinant D the principal diagonal is covered by two square matrices without elements in common, with determinants D_1 and D_2, and to one side of them all the elements are equal to zero, then $D = D_1 D_2$.

EXAMPLE 3.

$$\begin{vmatrix} 3 & 2 & 0 & 0 \\ 4 & 5 & 0 & 0 \\ 6 & 7 & 2 & 3 \\ 8 & 9 & 3 & 7 \end{vmatrix} = \begin{vmatrix} 3 & 2 \\ 4 & 5 \end{vmatrix} \times \begin{vmatrix} 2 & 3 \\ 3 & 7 \end{vmatrix} = 7 \times 5 = 35.$$

A more general case is that of the so-called *echelon* determinants: if in the determinant D there stand along the principal diagonal a chain of square matrices with determinants D_1, D_2, ..., D_k, and on one side of this chain all the elements are equal to zero, then $D = D_1 D_2 \dots D_k$.

EXAMPLE 4.

$$\begin{vmatrix} 1 & 2 & 0 & 0 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 & 0 & 0 \\ 5 & 2 & 3 & 0 & 0 & 0 & 0 \\ 7 & 6 & 2 & 5 & 4 & 0 & 0 \\ 3 & 2 & 7 & 4 & 5 & 0 & 0 \\ 5 & 1 & 3 & 2 & 6 & 7 & 3 \\ 2 & 7 & 1 & 5 & 3 & 4 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} \times 3 \times \begin{vmatrix} 5 & 4 \\ 4 & 5 \end{vmatrix} \times \begin{vmatrix} 7 & 3 \\ 4 & 1 \end{vmatrix}$$
$$= (-2) \times 3 \times 9 \times (-5) = 270.$$

10. The expansion of a determinant with respect to a row and a column. Bordered determinants.

A *bordered determinant for the determinant*

$$D = \begin{vmatrix} a_{11} \dots a_{1n} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_{n1} \dots a_{nn} \end{vmatrix}$$

is a determinant

$$D_1 = \begin{vmatrix} a_{11} \dots a_{1n} & x_1 \\ \cdot \quad \cdot \quad \cdot \quad \cdot & \cdot \\ a_{n1} \dots a_{nn} & x_n \\ y_1 \dots y_n & z \end{vmatrix}.$$

Any determinant of order $n > 1$ can be regarded as bordered with respect to the minor M_{nn} of the element standing in its lower right-hand corner. By expanding D_1 with respect to its last column, and then all the minors obtained, except the

minor for z, with respect to the last row, we obtain an expansion of this determinant with respect to the last column and the last row:

$$D_1 = Dz - \sum_{i,j=1}^{n} A_{ij}x_i y_j ,$$

where A_{ij} is the algebraic complement (cofactor) of the element a_{ij} in the determinant D (see [30], p. 22, or [23], p. 61).

If it is required to expand a determinant with respect to any row and any column, then, by interchanging the given row and column with all those following them, we arrive at the case considered.

Expansion with respect to a row and a column can be applied conveniently in the case when it is easy to evaluate the cofactors of all the elements of the determinant D.

EXAMPLE 1. To evaluate the determinant

$$D_1 = \begin{vmatrix} a_1 & 0 & \dots & 0 & x_1 \\ 0 & a_2 & \dots & 0 & x_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & a_n & x_n \\ y_1 & y_2 & \dots & y_n & z \end{vmatrix}.$$

Here

$$D = \begin{vmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & a_n \end{vmatrix},$$

$A_{ij} = 0$ when $i \neq j$, $A_{ii} = D/a_i$. Therefore

$$D_1 = D\left(z - \sum_{i=1}^{n} \frac{x_i y_i}{a_i}\right)$$

where $D = a_1 a_2 \dots a_n$.

We consider the bordered determinant

$$\Delta = \begin{vmatrix} a_{11} & a_{12} \dots a_{1n} & x_{11} \dots x_{1p} \\ a_{21} & a_{22} \dots a_{2n} & x_{21} \dots x_{2p} \\ \cdot & \cdot \cdot \cdot \cdot & \cdot \cdot \cdot \cdot \\ a_{n1} & a_{n2} \dots a_{nn} & x_{n1} \dots x_{np} \\ y_{11} & y_{12} \dots y_{1n} & 0 \dots 0 \\ \cdot & \cdot \cdot \cdot \cdot & \cdot \cdot \cdot \cdot \\ y_{p1} & y_{p2} \dots y_{pn} & 0 \dots 0 \end{vmatrix}.$$

Expanding this determinant by Laplace's theorem with respect to the last p columns, and then (when $n \geqslant p$) each of the complementary minors of order n with respect to the last p rows, we obtain:

(1) If $p \leqslant n$, then

$$\Delta = (-1)^p \sum A_{i_1 i_2 \cdots i_p j_1 j_2 \cdots j_p} X_{i_1 i_2 \cdots i_p} Y_{j_1 j_2 \cdots j_p},$$

where $A_{i_1 i_2 \cdots i_p j_1 j_2 \cdots j_p}$ is the algebraic complement of the minor $M_{i_1 i_2 \cdots i_p j_1 j_2 \cdots j_p}$ in the determinant $D = |a_{ij}|_1^n$,

$$X_{i_1 i_2 \cdots i_p} = \begin{vmatrix} x_{i_1 1} & x_{i_1 2} & \cdots & x_{i_1 p} \\ x_{i_2 1} & x_{i_2 2} & \cdots & x_{i_2 p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{i_p 1} & x_{i_p 2} & \cdots & x_{i_p p} \end{vmatrix},$$

$$Y_{j_1 j_2 \cdots j_p} = \begin{vmatrix} y_{1 j_1} & y_{1 j_2} & \cdots & y_{1 j_p} \\ y_{2 j_1} & y_{2 j_2} & \cdots & y_{2 j_p} \\ \cdots & \cdots & \cdots & \cdots \\ y_{p j_1} & y_{p j_2} & \cdots & y_{p j_p} \end{vmatrix}$$

and the sum is taken with respect to all the indices $i_1, i_2, \ldots, i_p$; $j_1, j_2, \ldots, j_p$, satisfying the conditions

$$1 \leqslant \begin{matrix} i_1 < i_2 < \ldots < i_p \\ j_1 < j_2 < \cdots < j_p \end{matrix} \leqslant n.$$

The same sort of expansion can be given for a bordered determinant D by means of the first p rows and the first p columns.

(2) If $p > n$, then $\Delta = 0$.

EXAMPLE 2.

$$\Delta = \begin{vmatrix} a_1 & 0 & 0 & x_{11} & x_{12} \\ 0 & a_2 & 0 & x_{21} & x_{22} \\ 0 & 0 & a_3 & x_{31} & x_{32} \\ y_{11} & y_{12} & y_{13} & 0 & 0 \\ y_{21} & y_{22} & y_{23} & 0 & 0 \end{vmatrix}$$

$$= a_1 \begin{vmatrix} x_{21} & x_{22} \\ x_{31} & x_{32} \end{vmatrix} \begin{vmatrix} y_{12} & y_{13} \\ y_{22} & y_{23} \end{vmatrix} + a_2 \begin{vmatrix} x_{11} & x_{12} \\ x_{31} & x_{32} \end{vmatrix} \begin{vmatrix} y_{11} & y_{13} \\ y_{21} & y_{23} \end{vmatrix}$$

$$+ a_3 \begin{vmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{vmatrix} \begin{vmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{vmatrix}.$$

11. The multiplication of determinants. The Binet–Cauchy theorem.

The *product of the determinants* $|\mathbf{A}| = |a_{ij}|_1^n$ and $|\mathbf{B}| = |b_{ij}|_1^n$ of the same order n is determinant $|\mathbf{C}| = |c_{ij}|_1^n$, of the same order, all of whose elements can be calculated by means of any of the following four formulae:

$$(1)\quad c_{ij} = a_{i1}b_{j1} + a_{i2}b_{j2} + \ldots + a_{in}b_{jn};$$
$$(2)\quad c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj};$$
$$(3)\quad c_{ij} = a_{1i}b_{j1} + a_{2i}b_{j2} + \ldots + a_{ni}b_{jn};$$
$$(4)\quad c_{ij} = a_{1i}b_{1j} + a_{2i}b_{2j} + \ldots + a_{ni}b_{nj}.$$

In case (1) the element c_{ij} is obtained as the sum of products of the elements of the ith row of the determinant $|\mathbf{A}|$ by the corresponding elements of the jth row of the determinant $|\mathbf{B}|$. Therefore we say that in the given case the product is obtained by mutiplying the rows of the first determinant by the rows of the second. In the second case it is obtained by multiplying rows by columns, in the third by multiplying columns by rows, and in the fourth by multiplying columns by columns. The value of the determinant $|\mathbf{C}|$ is the same in all four cases, although the elements c_{ij} are different: $|\mathbf{C}| = |\mathbf{A}| \times |\mathbf{B}|$. The multiplication of determinants is closely connected with the multiplication of matrices (see pp. 52–53).

EXAMPLE 1. Multiplying the determinants

$$\begin{vmatrix} 2 & 1 \\ 4 & 3 \end{vmatrix} = 2, \qquad \begin{vmatrix} 4 & -3 \\ -1 & 2 \end{vmatrix} = 5,$$

together in the four ways, we obtain:

$$(1)\ \begin{vmatrix} 5 & 0 \\ 7 & 2 \end{vmatrix} = 10,\ (2)\ \begin{vmatrix} 7 & -4 \\ 13 & -6 \end{vmatrix} = 10,$$

$$(3)\ \begin{vmatrix} -4 & 6 \\ -5 & 5 \end{vmatrix} = 10,\ (4)\ \begin{vmatrix} 4 & 2 \\ 1 & 3 \end{vmatrix} = 10.$$

The multiplication of determinants can be used to evaluate certain determinants of special form.

EXAMPLE 2. To evaluate the determinant

$$D_x = \begin{vmatrix} x_1 & x_2 & x_3 & x_4 \\ -x_2 & x_1 & x_4 & -x_3 \\ -x_3 & -x_4 & x_1 & x_2 \\ -x_4 & x_3 & -x_2 & x_1 \end{vmatrix}.$$

Squaring it by multiplying rows by rows, we obtain along the principal diagonal the same expression $x_1^2 + x_2^2 + x_3^2 + x_4^2$, and zeros outside the principal diagonal. Therefore $D_x^2 = (x_1^2 + x_2^2 + x_3^2 + x_4^2)^4$. Because D_x contains

the product x_1^4 along the principal diagonal, we can take square roots with the plus sign on both sides of the least equation, whence $D_x = (x_1^2 + x_2^2 + x_3^2 + x_4^2)^2$.

EXAMPLE 3. To evaluate the determinant

$$D = \begin{vmatrix} s_0 & s_1 & s_2 & \dots & s_{n-1} \\ s_1 & s_2 & s_3 & \dots & s_n \\ s_2 & s_3 & s_4 & \dots & s_{n+1} \\ \dots & \dots & \dots & \dots & \dots \\ s_{n-1} & s_n & s_{n+1} & \dots & s_{2n-2} \end{vmatrix},$$

where $s_k = x_1^k + x_2^k + \dots + x_n^k$ $(k = 0, 1, 2, \dots)$ — a so-called *power sum* of the variables $x_1, x_2, \dots, x_n$ (in particular, $s_0 = n$).

On multiplying the Vandermonde determinant

$$V_n = \begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} \end{vmatrix}$$

by itself by means of the multiplication of columns by columns, and applying the expression for a Vandermonde determinant given on p. 13, we obtain

$$D = V_n^2 = \prod_{n \geqslant i > j \geqslant 1} (x_i - x_j)^2 .$$

The Binet–Cauchy theorem. Suppose that we are given two rectangular matrices

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \dots a_{1n} \\ a_{21} & a_{22} \dots a_{2n} \\ \dots & \dots \\ a_{p1} & a_{p2} \dots a_{pn} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \dots b_{1n} \\ b_{21} & b_{22} \dots b_{2n} \\ \dots & \dots \\ b_{p1} & b_{p2} \dots b_{pn} \end{pmatrix}$$

of the same dimensions $p \times n$. Multiplying the rows of $\mathbf{A}$ by the rows of $\mathbf{B}$, we form the p^2 numbers

$$c_{ij} = a_{i1}b_{j1} + a_{i2}b_{j2} + \dots + a_{in}b_{jn} \qquad (i, j = 1, 2, \dots, p)$$

and we consider the determinant $D = |c_{ij}|_1^p$ of order p of these numbers.

Then: (1) If $p \leqslant n$, then

$$D = \sum_{1 \leqslant i_1 < i_2 < \dots < i_p \leqslant n} \mathbf{A}\,(i_1 i_2 \dots i_p)\,\mathbf{B}\,(i_1 i_2 \dots i_p),$$

where $\mathbf{A}(i_1 i_2 \dots i_p)$ and $\mathbf{B}(i_1 i_2 \dots i_p)$ are the minors of order p of the matrices $\mathbf{A}$ and $\mathbf{B}$ respectively, formed from the columns with the numbers $i_1, i_2, \dots, i_p$.

(2) If $p > n$, then $D = 0$ (see [6], p. 17).

The rule for the multiplication of determinants is a special case of the Binet–Cauchy theorem when $p = n$.

EXAMPLE 4. By applying the Binet–Cauchy theorem to the matrices

$$\begin{pmatrix} a_1 a_2 \dots a_n \\ b_1 b_2 \dots b_n \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} c_1 c_2 \dots c_n \\ d_1 d_2 \dots d_n \end{pmatrix}$$

we obtain the equation

$$\begin{vmatrix} a_1c_1 + a_2c_2 + \dots + a_nc_n & a_1d_1 + a_2d_2 + \dots + a_nd_n \\ b_1c_1 + b_2c_2 + \dots + b_nc_n & b_1d_1 + b_2d_2 + \dots + b_nd_n \end{vmatrix} = \sum_{1 \leqslant i < k \leqslant n} \begin{vmatrix} a_i & a_k \\ b_i & b_k \end{vmatrix} \times \begin{vmatrix} c_i & c_k \\ d_i & d_k \end{vmatrix}.$$

12. Circulants and the determinants connected with them.

A *circulant* (or circulant determinant) is either of the two following determinants:

$$D_n = \begin{vmatrix} a_1 & a_2 & a_3 \dots a_{n-1} & a_n \\ a_n & a_1 & a_2 \dots a_{n-2} & a_{n-1} \\ a_{n-1} & a_n & a_1 \dots a_{n-3} & a_{n-2} \\ \cdot & \cdot & \cdot \quad \cdot \quad \cdot & \cdot \\ a_2 & a_3 & a_4 \dots a_n & a_1 \end{vmatrix},$$

$$D'_n = \begin{vmatrix} a_1 & a_2 & a_3 \dots a_{n-1} & a_n \\ a_2 & a_3 & a_4 \dots a_n & a_1 \\ a_3 & a_4 & a_5 \dots a_1 & a_2 \\ \cdot & \cdot & \cdot \quad \cdot \quad \cdot & \cdot \\ a_n & a_1 & a_2 \dots a_{n-2} & a_{n-1} \end{vmatrix}.$$

For the first of these we have the expression

$$D_n = f(\varepsilon_0) f(\varepsilon_1) f(\varepsilon_2) \dots f(\varepsilon_{n-1}),$$

where

$$f(x) = a_1 + a_2x + a_3x^2 + \dots + a_nx^{n-1}$$

and

$$\varepsilon_k = \cos\frac{2\pi k}{n} + i \sin\frac{2\pi k}{n}, \qquad k = 0, 1, 2, \dots, n-1,$$

are all the values of $\sqrt[n]{1}$ (see [30], p. 36, or [20], Problem 479).

By reversing the order of the rows with the numbers $2, 3, \ldots, n$ in the determinant D_n we obtain an expression for the second determinant:

$$D_n' = (-1)^{(n-1)(n-2)/2} D_n .$$

EXAMPLE. To evaluate the determinant

$$D = \begin{vmatrix} 1 & a & a^2 & \ldots & a^{n-1} \\ a^{n-1} & 1 & a & \ldots & a^{n-2} \\ a^{n-2} & a^{n-1} & 1 & \ldots & a^{n-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a & a^2 & a^3 & \ldots & 1 \end{vmatrix} .$$

Here

$$f(x) = 1 + ax + a^2x^2 + \ldots + a^{n-1}x^{n-1} = \frac{1 - a^n x^n}{1 - ax} .$$

First let $a \neq 0$ and $a^n \neq 1$. Because $\varepsilon_i^n = 1$, $i = 0, 1, 2, \ldots, n-1$, then

$$D = f(\varepsilon_0) f(\varepsilon_1) \ldots f(\varepsilon_{n-1}) = \frac{(1 - a^n)^n}{(1 - a\varepsilon_0)(1 - a\varepsilon_1) \ldots (1 - a\varepsilon_{n-1})} .$$

To simplify the denominator we set $x = 1/a$ in the equation

$$x^n - 1 = (x - \varepsilon_0)(x - \varepsilon_1) \ldots (x - \varepsilon_{n-1}) .$$

We obtain

$$\frac{1}{a^n} - 1 = \left(\frac{1}{a} - \varepsilon_0\right)\left(\frac{1}{a} - \varepsilon_1\right) \ldots \left(\frac{1}{a} - \varepsilon_{n-1}\right)$$

or, multiplying by a^n, we find

$$1 - a^n = (1 - a\varepsilon_0)(1 - a\varepsilon_1) \ldots (1 - a\varepsilon_{n-1}).$$

This can also be proved in another way, by expanding the brackets on the right-hand side and applying Vieta's formula for $x^n - 1$. Hence,

$$D_n = (1 - a^n)^{n-1} .$$

This formula is also true when $a = 0$, because both its sides reduce to one, and when $a^n = 1$, because with this hypothesis, on multiplying the second row of D by a we obtain a row which coincides with the first, and therefore $D = 0$.

A *skew circulant* is a determinant of the form

$$Z_n = \begin{vmatrix} a_1 & a_2 & a_3 \ldots a_n \\ -a_n & a_1 & a_2 \ldots a_{n-1} \\ -a_{n-1} & -a_n & a_1 \ldots a_{n-2} \\ \cdot & \cdot & \cdot \\ -a_2 & -a_3 & -a_4 \ldots a_1 \end{vmatrix} .$$

For this the expression

$$Z_n = f(\eta_0) f(\eta_1) f(\eta_2) \ldots f(\eta_{n-1}) ,$$

holds, where

$$f(x) = a_1 + a_2 x + a_3 x^2 + \ldots + a_n x^{n-1}$$

and

$$\eta_k = \cos \frac{(2k+1)\pi}{n} + i \sin \frac{(2k+1)\pi}{n}, \quad k = 0, 1, 2, \ldots, n-1,$$

are all the values of the root $\sqrt[n]{-1}$ (see [20], Problem 493). Both the circulant and skew-circulant are special cases of the determinant where $z = r(\cos\varphi + i \sin\varphi)$ is any complex number and

$$\begin{vmatrix} a_1 & a_2 & a_3 & \ldots & a_n \\ a_n z & a_1 & a_2 & \ldots & a_{n-1} \\ a_{n-1} z & a_n z & a_1 & \ldots & a_{n-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_2 z & a_3 z & a_4 z & \ldots & a_1 \end{vmatrix} = f(\alpha_0) f(\alpha_1) f(\alpha_2) \ldots f(\alpha_{n-1}),$$

are all the values of the root $\sqrt[n]{z}$.

$$a_k = \sqrt[n]{r}\left(\cos \frac{\varphi + 2k\pi}{n} + i \sin \frac{\varphi + 2k\pi}{n}\right), \quad k = 0, 1, 2, \ldots, n-1,$$

Expressions for various circulants of special forms are given in [20], Problems 375, 376, 484–92, 495.

13. Adjoint (or reciprocal) and associated determinants, their minors and Sylvester's formula.

There is no firmly established notation and terminology here.

The *adjoint* (or *reciprocal*) determinant for the determinant $D = |a_{ij}|_1^n$ of order $n \geqslant 2$ is the determinant $\hat{D} = |\hat{a}_{ij}|_1^n$, where $\hat{a}_{ij} = A_{ij} (i, j = 1, 2, \ldots, n)$, i.e. the determinant obtained from the given determinant D by replacing its elements by their cofactors (while preserving their order). The values of the given determinant and its adjoint are connected by the equation $\hat{D} = D^{n-1}$, where n is the order of D and $\hat{D}$ (see [20], Problem 506). In particular, if $D = 0$, then $\hat{D} = 0$ also, and conversely.

The minors of the given determinant and its adjoint are connected in the following way. Let M be a minor of order p of the determinant D of order n, A the algebraic complement

of the minor M, $\hat{M}$ the minor of the adjoint determinant $\hat{D}$ corresponding to the minor M, i.e. standing in the same rows and columns as M. Then $\hat{M} = D^{p-1}A$ (see [3], p. 37).

Denoting by $\hat{A}$ the algebraic complement of the minor $\hat{M}$ and applying the last equation with p replaced by $n-p$, we obtain $\hat{A} = D^{n-p-1}M$.

If M is a minor of the second order, standing in the ith and jth rows and the kth and lth columns of the determinant D, and M' is the complementary minor obtained from the determinant D by crossing out the given rows and columns, then

$$\begin{vmatrix} A_{ik}A_{il} \\ A_{jk}A_{jl} \end{vmatrix} = (-1)^{i+j+k+l}DM',$$

where A_{st} is the cofactor of the element a_{st} (see [20], Problem 508, or [3], p. 39).

From the last equation it follows that if the given determinant D is equal to zero, then all the minors of the second order of the adjoint determinant $\hat{D}$ are equal to zero, i.e. any two rows (or any two columns) of the determinant $\hat{D}$ are proportional.

Denoting by $\hat{M}_{ij}$ the minor and by $\hat{A}_{ij}$ the cofactor of the element $\hat{a}_{ij}$ of the adjoint determinant $\hat{D}$, we obtain

$$\hat{M}_{ij} = (-1)^{i+j}D^{n-2}a_{ij}$$

or

$$\hat{A}_{ij} = D^{n-2}a_{ij}$$

(see [3], p. 39).

This equation enables us to find the elements of the determinant D itself by means of the elements of the adjoint determinant $\hat{D}$ (or by means of the minors of the determinant D) (see [20], Problem 512).

The *associated* determinant for the determinant $D = |a_{ij}|_1^n$ ($n \geqslant 2$) is the determinant $\tilde{D} = |\tilde{a}_{ij}|_1^n$, where $\tilde{a}_{ij} = M_{ij}$ ($i, j = 1, 2, \ldots, n$), i.e. the determinant obtained from the given determinant D by replacing its elements by their minors. The properties of this determinant are similar to the properties of the adjoint determinant $\hat{D}$ and can be derived from the latter. In fact: $\tilde{D} = \hat{D} = D^{n-1}$. If M is a minor of order p of the determinant D, M' is the complementary minor for M and $\tilde{M}$ is the minor of the determinant $\tilde{D}$ associated with D

which stands in the same rows and columns as the minor M, then $\tilde{M} = D^{p-1}M'$ (see [7], p. 19).

The concept of adjoint and associated determinants can be generalized by taking determinants consisting of minors of any order k of the determinant D or their algebraic complements. Let D be a determinant of order $n \geqslant 2$ and let k be equal to one of the numbers $1, 2, \ldots, n$. Let

$$N = C_n^k = \frac{n!}{k!(n-k)!}.$$

We number all the sets of k numbers from $1, 2, \ldots, n$ in an arbitrary (but subsequently fixed) order: $s_1, s_2, \ldots, s_N$. Let μ_{ij} be the minor of order k standing in the rows of the determinant D whose numbers form the set s_i and in the columns whose numbers form the set s_j, and let α_{ij} be the algebraic complement of the minor μ_{ij} in D. Then the *k-th associated determinant* of the determinant D is the determinant $\tilde{D}_k = |\mu_{ij}|_1^N$.

The *k-th adjoint* (or the *k-th reciprocal*) of the determinant D is the determinant $\hat{D}_k = |\alpha_{ij}|_1^N$. $\tilde{D}_k$ and $\hat{D}_k$ have order $N = \binom{k}{n}$. These determinants have the following properties (see [20], Problem 551):

(1) the values of the determinants $\tilde{D}_k$ and $\hat{D}_k$ are not altered by the choice of the numbering of the sets of k numbers from the numbers $1, 2, \ldots, n$;

$$(2) \qquad \tilde{D}_k = \hat{D}_{n-k};$$

$$(3) \qquad \tilde{D}_k \hat{D}_k = D^N = D^{C_n^k}$$

$$(4) \qquad \tilde{D}_k = D^{C_{n-1}^{k-1}};$$

$$(5) \qquad \hat{D}_k = D^{C_{n-1}^k}.$$

Sylvester's theorem. Let $D = |a_{ij}|_1^n$ be a determinant of order $n \geqslant 2$, $M = |a_{ij}|_1^p$ be the minor of order p $(1 \leqslant p < n)$ standing in its upper left-hand corner, b_{ij} the minor obtained by adjoining below the minor M the ith row of the determinant D and to the right of it the jth column of the determinant $D(i, j = p+1, p+2, \ldots, n)$, and $\Delta = |b_{ij}|_{p+1}^n$ the determinant of order $n-p$ of all the elements b_{ij}. Then *Sylvester's formula*

$$\Delta = M^{n-p-1}D$$

holds, (see [7], p. 21).

14. Symmetric, skew-symmetric and pseudo-symmetric determinants.

The determinant $D = |a_{ij}|_1^n$ is called *symmetric*, *skew-symmetric* or *pseudo-symmetric* if its elements are connected, respectively, by the conditions:

$$(1)\ a_{ij} = a_{ji}, \qquad i, j = 1, 2, \ldots, n;$$
$$(2)\ a_{ij} = -a_{ji}, \quad i, j = 1, 2, \ldots, n;$$
$$(3)\ a_{ij} = -a_{ji}, \quad i, j = 1, 2, \ldots, n;\ i \neq j.$$

The principal minors (i.e. the minors which lie in a symmetric position with respect to the principal diagonal, i.e. which lie in rows and columns with the same numbers) of a determinant of one of these three types are themselves of the same type.

In a symmetric determinant the elements a_{ji} and a_{ij} which lie in symmetric positions with respect to the principal diagonal have equal minors and equal cofactors:

$$M_{ij} = M_{ji}, \qquad A_{ij} = A_{ji}.$$

In other words, the adjoint determinant $\hat{D}$ (p. 25) and the associated determinant $\tilde{D}$ (p. 26) of a symmetric determinant D will also be symmetric. The same is also true for the kth adjoint and the kth associated determinants $\hat{D}_k$ and $\tilde{D}_k$ (p.27).

If D is a skew-symmetric determinant of order n, then the adjoint determinant $\hat{D}$ and the associated determinant $\tilde{D}$ will be symmetric if n is odd, and skew-symmetric if n is even, and, in general, the kth adjoint determinant $\hat{D}_k$ will be symmetric if the difference $n - k$ is even, and skew-symmetric if it is odd, and the kth associated determinant $\tilde{D}_k$ will be symmetric if k is even, and skew-symmetric if k is odd.

A skew-symmetric determinant of odd order is zero (see [11], p. 43). A skew-symmetric determinant of even order n is equal to the square of a homogeneous polynomial of degree $n/2$ in its elements (see [30], p. 45, or [20], Problem 543). Because of the condition $a_{ij} = -a_{ij}$ this polynomial can be expressed simply in terms of the elements of the determinant which lie above the principal diagonal. Then it will be determined uniquely to within its sign and it is called a *Pfaffian aggregate* or *Pfaffian*. For an exact definition of a Pfaffian aggregate see [10], pp. 147–56, or [20], Problem 545.

Let $D_n = |a_{ij}|_1^n$ be a skew-symmetric determinant of even order n and let p_n be the Pfaffian aggregate corresponding to

it, so that $D_n = p_n^2$. We denote by p_{in} the Pfaffian aggregate of the skew-symmetric determinant D_{in}, obtained from the determinant D_n by crossing out the nth and ith rows, and also the nth and ith columns. Then the recurrence formula:

$$p_2 = a_{12}, \quad p_n = \sum_{i=1}^{n-1} (-1)^{i-1} p_{in} a_{in} \quad (n \geqslant 4)$$

holds (see [20], Problem 546).

EXAMPLE 1. $p_2 = a_{12}$. Therefore $p_{14} = a_{23}$, $p_{24} = a_{13}$, $p_{34} = a_{12}$ and by the recurrence formula

$$p_4 = p_{14}a_{14} - p_{24}a_{24} + p_{34}a_{34} = a_{23}a_{14} - a_{13}a_{24} + a_{12}a_{34}.$$

In addition,

$$\begin{aligned} p_6 = {} & p_{16}a_{16} - p_{26}a_{26} + p_{36}a_{36} - p_{46}a_{46} + p_{56}a_{56} \\ = {} & a_{34}a_{25}a_{16} - a_{24}a_{35}a_{16} + a_{23}a_{45}a_{16} - a_{34}a_{15}a_{26} \\ & + a_{14}a_{35}a_{26} - a_{13}a_{45}a_{26} + a_{24}a_{15}a_{36} - a_{14}a_{25}a_{36} \\ & + a_{12}a_{45}a_{36} - a_{23}a_{15}a_{46} + a_{13}a_{25}a_{46} - a_{12}a_{35}a_{46} + a_{23}a_{14}a_{56} \\ & - a_{13}a_{24}a_{56} + a_{12}a_{34}a_{56}. \end{aligned}$$

From the recurrence formula it is clear that the number of summands of the Pfaffian aggregate p_n is equal to the product $1 \times 3 \times 5 \ldots (n-1)$ (n is an even number).

A pseudo-symmetric determinant, in which all the elements of the principal diagonal are equal to each other, is a sum of squares of polynomials in its elements if it is of even order, and to a sum of the given form, multiplied by an element of the principal diagonal, if it is of odd order (see [30], p. 46).

EXAMPLE 2.

$$\begin{vmatrix} x & r & q \\ -r & x & p \\ -q & -p & x \end{vmatrix} = x(x^2 + p^2 + q^2 + r^2);$$

$$\begin{vmatrix} x & a & b & c \\ -a & x & r & q \\ -b & -r & x & p \\ -c & -q & -p & x \end{vmatrix} = x^4 + (a^2 + b^2 + c^2 + p^2 + q^2 + r^2)x^2 + (ap - bq + cr)^2.$$

15. Kronecker product of determinants.

For the Kronecker product of matrices, see pp. 101–103.

The *Kronecker* or direct product of two determinants $|\mathbf{A}| = |a_{ij}|_1^n$ of order n and $|\mathbf{B}| = |b_{ij}|_1^p$ of order p is the determinant of order np, whose matrix is constructed from the matrices **A** and **B** of the given determinants in one of the following

two ways: (1) this matrix consists of p^2 cells, arranged in p rows and p columns, where the cell in the ith row and jth columns of cells is obtained from the matrix $\mathbf{A}$ by multiplying all its elements by the number b_{ij}, or (2) this matrix consists of n^2 cells, arranged in n rows and n columns, where the cell in the ith row and jth column of cells is obtained from the matrix $\mathbf{B}$ by multiplying all its elements by the number a_{ij}. In the first case the product is said to be *left*, and in the second *right*. In both cases the Kronecker product has the same value $|\mathbf{A}|^p\,|\mathbf{B}|^n$.

EXAMPLE. The left product of the determinants

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \text{ and } |\mathbf{B}| = \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix}$$

has the form

$$\begin{vmatrix} a_{11}b_{11} & a_{12}b_{11} & a_{13}b_{11} & a_{11}b_{12} & a_{12}b_{12} & a_{13}b_{12} \\ a_{21}b_{11} & a_{22}b_{11} & a_{23}b_{11} & a_{21}b_{12} & a_{22}b_{12} & a_{23}b_{12} \\ a_{31}b_{11} & a_{32}b_{11} & a_{33}b_{11} & a_{31}b_{12} & a_{32}b_{12} & a_{33}b_{12} \\ a_{11}b_{21} & a_{12}b_{21} & a_{13}b_{21} & a_{11}b_{22} & a_{12}b_{22} & a_{13}b_{22} \\ a_{21}b_{21} & a_{22}b_{21} & a_{23}b_{21} & a_{21}b_{22} & a_{22}b_{22} & a_{23}b_{22} \\ a_{31}b_{21} & a_{32}b_{21} & a_{33}b_{21} & a_{31}b_{22} & a_{32}b_{22} & a_{33}b_{22} \end{vmatrix} = |\mathbf{A}|^2\,|\mathbf{B}|^3.$$

16. Hadamard's inequality and other inequalities connected with determinants.

Let $\mathbf{A} = (a_{ij})_1^n$ be a real square matrix. For its determinant we have *Hadamard's inequality*:

$$|\mathbf{A}|^2 \leqslant \prod_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2.$$

If $|\mathbf{A}| \neq 0$, then Hadamard's inequality reduces to an equality if and only if the rows of the matrix $\mathbf{A}$ are orthogonal in pairs, i.e. when the equation

$$\sum_{k=1}^{n} a_{ik}a_{jk} = 0; \quad i, j = 1, 2, \ldots, n; \quad i \neq j$$

is satisfied (see [7], p. 51).

Hadamard's inequality has the following geometrical meaning. If the elements of the rows of the matrix $\mathbf{A}$ are regarded as the coordinates of n-dimensional vectors of Euclidean space ([11], pp. 211–2) with an ortho-normal basis,

then the determinant $|\mathbf{A}|$, to within sign, is equal to the volume of the n-dimensional parallelepiped constructed on these vectors as edges. Hadamard's inequality means that the volume of the parallelepiped does not exceed the product of the lengths of its edges which meet at one vertex, and is equal to this product if and only if the edges are orthogonal in pairs.

We call the real square matrix $\mathbf{A}$ *positive definite* if all its minors which stand in the upper left-hand corner are positive, i.e. if the *quadratic form* with matrix $\mathbf{A}$ is positive definite. For such a matrix Hadamard's inequality can be written in another form.

Second form of Hadamard's inequality. If $\mathbf{A} = (a_{ij})_1^n$ is a positive definite matrix, then

$$|\mathbf{A}| \leqslant a_{11}a_{22} \dots a_{nn},$$

where the equal sign occurs if and only if the matrix $\mathbf{A}$ is diagonal, i.e. when $a_{ij} = 0$; $i \neq j$; $i, j = 1, 2, \dots, n$ ([7], p. 48).

The determinant formed from the elements of the matrix $\mathbf{A}$ standing at the intersections of the k rows with the numbers $i_1, i_2, \dots, i_k$ and the k columns with the numbers $j_1, j_2, \dots, j_k$, is called a *minor of the k-th order* of the matrix $\mathbf{A}$ and is denoted by

$$\mathbf{A}\begin{pmatrix} i_1 i_2 \dots i_k \\ j_1 j_2 \dots j_k \end{pmatrix}.$$

Generalized form of Hadamard's inequality. If $\mathbf{A} = (a_{ij})_1^n$ is a positive definite matrix and p is a natural number $(1 \leqslant p \leqslant n-1)$, then

$$|\mathbf{A}| \leqslant \mathbf{A}\begin{pmatrix} 1 & 2 \dots p \\ 1 & 2 \dots p \end{pmatrix} \mathbf{A}\begin{pmatrix} p+1 \dots n \\ p+1 \dots n \end{pmatrix}, \tag{1.1}$$

where the equal sign occurs if and only if $a_{ij} = a_{ji} = 0$; $i = 1, 2, \dots, p$; $j = p+1, \dots, n$ (see [7], p. 50).

If we make use of the multiplication of rectangular matrices (pp. 52–53) and denote by $\mathbf{A}'$ the transpose of the matrix $\mathbf{A}$, then the generalized form of Hadamard's inequality can be written in the following matrix form.

Let $\mathbf{A} = (\mathbf{B}, \mathbf{C})$ be a real square matrix, where $\mathbf{B}$ is the matrix from the first p, $\mathbf{C}$ the matrix from the last $n-p$ columns of the matrix $\mathbf{A}$ $(1 \leqslant p \leqslant n-1)$. Then

$$|\mathbf{A}'\mathbf{A}| \leqslant |\mathbf{B}'\mathbf{B}| \times |\mathbf{C}'\mathbf{C}|.$$

If $\mathbf{B}'\mathbf{C} = \mathbf{0}$, then the equal sign occurs. If $|\mathbf{A}| \neq \mathbf{0}$, then, conversely, when the equal sign occurs we obtain $\mathbf{B}'\mathbf{C} = \mathbf{0}$ (the part of the matrix $\mathbf{A}$ in (1.1) is now played by the matrix $\mathbf{A}'\mathbf{A}$). In this form the inequality has the following generalization.

Let $\mathbf{A} = (\mathbf{B}, \mathbf{C})$ be a real rectangular matrix, where $\mathbf{B}$ and $\mathbf{C}$ are respectively matrices from some of the first columns and from the remaining columns of the matrix $\mathbf{A}$. Then

$$|\mathbf{A}'\mathbf{A}| \leqslant |\mathbf{B}'\mathbf{B}|\,|\mathbf{C}'\mathbf{C}|.$$

If $\mathbf{B}'\mathbf{C} = \mathbf{0}$, then the equality sign occurs. If the columns of the matrix $\mathbf{A}$ (as vectors) are linearly independent ([34], pp. 56–71), then, conversely, when the equality sign occurs we obtain $\mathbf{B}'\mathbf{C} = \mathbf{0}$ ([20], Problem 922, or [24], Problem 518).

An analogous inequality is also true for matrices with complex elements. We denote by $\mathbf{A}^*$ the matrix which is obtained from the matrix $\mathbf{A}$ by transposing and replacing all the elements by their complex conjugates. Let $\mathbf{A} = (\mathbf{B}, \mathbf{C})$ be a complex rectangular matrix, where $\mathbf{B}$ and $\mathbf{C}$ are matrices respectively from some of the first and the remaining columns of the matrix $\mathbf{A}$. Then the determinants

$$|\mathbf{A}^*\mathbf{A}|,\ |\mathbf{B}^*\mathbf{B}|,\ |\mathbf{C}^*\mathbf{C}|$$

will be non-negative real numbers, where

$$|\mathbf{A}^*\mathbf{A}| \leqslant |\mathbf{B}^*\mathbf{B}|\,|\mathbf{C}^*\mathbf{C}|.$$

If $\mathbf{B}^*\mathbf{C} = \mathbf{0}$, then the equality sign occurs. If the columns of the matrix $\mathbf{A}$ are linearly independent, then, conversely, if the equality sign occurs, the equation $\mathbf{B}^*\mathbf{C} = \mathbf{0}$ follows ([20], Problem 925, or [24], Problem 521).

We mention two further approximations for the value of a determinant.

If all the elements of the determinant $D = |a_{ij}|_1^n$ of order n have moduli which do not exceed the positive number M, then for the modulus of this determinant the approximation $|D| \leqslant M^n \sqrt{n^n}$ is true, and this approximation is exact ([24], Problems 522, 524, or [20], Problem 926).

If all the elements of the determinant $D = |a_{ij}|_1^n$ are real and satisfy the condition $0 \leqslant a_{ij} \leqslant M$, then

$$|D| \leqslant \frac{M^n \sqrt{[(n+1)^{n+1}]}}{2^n}$$

(see [24], Problem 523).

2. Systems of Linear Equations

1. General information.

A *system of linear equations* has the form

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2, \\ \cdot\ \cdot\ \cdot

first system, then the second system will be a consequence of the first. If the first system is consistent and the second system is a consequence of the first, then every equation of the second system will be a linear combination of equations of the first system.

2. Gauss's method of eliminating unknowns.

Elementary transformations of a system of linear equations are transformations of the following three types:

(1) interchange of two equations;
(2) multiplication of both sides of one of the equations by any number distinct from zero;
(3) addition to both sides of one equation of the corresponding sides of another, multiplied by any number.

Elementary transformations reduce a given system of equations to an equivalent system.

A *system of echelon form* (or an *echelon system*) is a system of the form

$$\begin{aligned}
&a_{1,k_1}x_{k_1} + a_{1,k_1+1}x_{k_1+1} + \ldots + a_{1n}x_n = b_1,\\
&a_{2,k_2}x_{k_2} + a_{2,k_2+1}x_{k_2+1} + \ldots + a_{2n}x_n = b_2,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&a_{r,k_r}x_{k_r} + a_{r,k_r+1}x_{k_r+1} + \ldots + a_{rn}x_n = b_r,\\
&0 = 1,\\
&0 = 0,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&0 = 0,
\end{aligned}$$

where $1 \leqslant k_1 < k_2 < \ldots < k_r \leqslant n$ and $a_{i,k_i} \neq 0$, $i = 1, 2, \ldots, r$. Here we denote by $0 = 1$ and $0 = 0$ equations in which the coefficients for all the unknowns are equal to zero. Here the equations of the form $0 = 1$ and $0 = 0$ may possibly not occur, but an equation of the form $0 = 1$ occurs not more than once.

EXAMPLE 1. The system

$$\begin{aligned}
x_1 + 2x_2 \qquad &+ x_4 - x_5 + \quad x_6 + 3x_7 = 5,\\
x_2 - x_3 \qquad &+ x_5 \qquad - \quad x_7 = 2,\\
&3x_4 + x_5 - \quad x_6 + \quad x_7 = 1,\\
&\quad x_5 + 2x_6 + 3x_7 = 0,\\
&\qquad\qquad 0 = 1,\\
&\qquad\qquad 0 = 0,
\end{aligned}$$

is of the echelon form.

Any system of linear equations can be reduced to echelon form by means of a finite number of elementary transformations. In this way the solution of any system can be reduced to the solution of a system in echelon form. Any equations of an echelon system of the form $0 = 0$ can be rejected, because this reduces to a system of equations equivalent to the previous one. An echelon system is consistent if and only if it does not contain an equation of the form $0 = 1$. The unknowns $x_{k_1}, x_{k_2}, \ldots, x_{k_r}$, which are the initial unknowns in the equations of an echelon system, are called the principal unknowns, and the remaining unknowns are said to be free.

The solution of a consistent system of echelon form is found thus: we reject equations of the form $0 = 0$, then from the last equation we find an expression for the last principal unknown in terms of the free unknowns and we substitute this expression in all the previous equations; in the last of the equations obtained we again find an expression for the principal unknowns in terms of the free unknowns and so on. The expressions for the principal unknowns in terms of the free unknowns obtained in this way are called the *general solution* of the system of equations. By substituting arbitrary numbers in the general solution instead of the free unknowns, we obtain completely determined values for the principal unknowns.

An echelon system containing free unknowns and not containing an equation of the form $0 = 1$ has an infinite number of solutions. If, however an echelon system does not contain free unknowns and does not contain an equation of the form $0 = 1$, then it has a unique solution.

An echelon system is called *triangular* if it does not contain any equations in which the coefficients for all the unknowns are equal to zero, and if the number of equations is equal to the number of unknowns. A system of linear equations has a unique solution if and only if it can be reduced to triangular form by means of elementary transformations and by rejecting equations of the form $0 = 0$.

In practice it is more convenient to reduce to echelon form not the system of equations itself, but the matrix of the coefficients of the unknowns and the free terms; we join the matrices obtained consecutively by the equivalence sign $\sim$.

EXAMPLE 2. To solve the system of equations

$$\begin{aligned} x_1 + x_2 \quad\quad + 2x_4 + 3x_5 &= 5, \\ 2x_1 + 4x_2 - x_3 + 5x_4 + 4x_5 &= -1, \end{aligned}$$

$$\begin{aligned} x_1 + 3x_2 \qquad\quad + 5x_4 + 2x_5 &= -3, \\ 3x_1 + 7x_2 - 3x_3 + 9x_4 + 2x_5 &= -14, \\ 2x_1 + 8x_2 - 4x_3 + 2x_4 + 7x_5 &= -10. \end{aligned}$$

We transform the matrix

$$\left(\begin{array}{rrrrr|r} 1 & 1 & 0 & 2 & 3 & 5 \\ 2 & 4 & -1 & 5 & 4 & -1 \\ 1 & 3 & 0 & 5 & 2 & -3 \\ 3 & 7 & -3 & 9 & 2 & -14 \\ 2 & 8 & -4 & 2 & 7 & -10 \end{array}\right)$$

(the first row, multiplied respectively by 2, 1, 3, 2, is subtracted from the remaining rows)

$$\sim \left(\begin{array}{rrrrr|r} 1 & 1 & 0 & 2 & 3 & 5 \\ 0 & 2 & -1 & 1 & -2 & -11 \\ 0 & 2 & 0 & 3 & -1 & -8 \\ 0 & 4 & -3 & 3 & -7 & -29 \\ 0 & 6 & -4 & -2 & 1 & -20 \end{array}\right)$$

(the second row, multiplied respectively by 1, 2, 3, is subtracted from the following rows)

$$\sim \left(\begin{array}{rrrrr|r} 1 & 1 & 0 & 2 & 3 & 5 \\ 0 & 2 & -1 & 1 & -2 & -11 \\ 0 & 0 & 1 & 2 & 1 & 3 \\ 0 & 0 & -1 & 1 & -3 & -7 \\ 0 & 0 & -1 & -5 & 7 & 13 \end{array}\right)$$

(the third row is added to the following rows)

$$\sim \left(\begin{array}{rrrrr|r} 1 & 1 & 0 & 2 & 3 & 5 \\ 0 & 2 & -1 & 1 & -2 & -11 \\ 0 & 0 & 1 & 2 & 1 & 3 \\ 0 & 0 & 0 & 3 & -2 & -4 \\ 0 & 0 & 0 & -3 & 8 & 16 \end{array}\right)$$

(the fourth row is added to the fifth row)

$$\sim \left(\begin{array}{rrrrr|r} 1 & 1 & 0 & 2 & 3 & 5 \\ 0 & 2 & -1 & 1 & -2 & -11 \\ 0 & 0 & 1 & 2 & 1 & 3 \\ 0 & 0 & 0 & 3 & -2 & -4 \\ 0 & 0 & 0 & 0 & 6 & 12 \end{array}\right).$$

The system of equations has been reduced to the triangular form

$$\begin{aligned} x_1 + x_2 \qquad\quad + 2x_4 + 3x_5 &= 5, \\ 2x_2 - x_3 + x_4 - 2x_5 &= -11, \\ x_3 + 2x_4 + x_5 &= 3, \\ 3x_4 - 2x_5 &= -4, \\ 6x_5 &= 12. \end{aligned}$$

It has a unique solution. From the last equation we find that $x_5 = 2$; we substitute this value for x_5 in the previous equation and from it we find that $x_4 = 0$; we substitute these values for x_4 and x_5 in the third equation and we find that $x_3 = 1$; in the same way we find that $x_2 = -3$, $x_1 = 2$.

EXAMPLE 3. To solve the system of equations

$$\begin{aligned} 2x_1 - x_2 - x_3 + 3x_4 &= 1, \\ 4x_1 - 2x_2 - x_3 + x_4 &= 5, \\ 6x_1 - 3x_2 - x_3 - x_4 &= 9, \\ 2x_1 - x_2 + 2x_3 - 12x_4 &= 10. \end{aligned}$$

We transform the matrix

$$\left(\begin{array}{cccc|c} 2 & -1 & -1 & 3 & 1 \\ 4 & -2 & -1 & 1 & 5 \\ 6 & -3 & -1 & -1 & 9 \\ 2 & -1 & 2 & -12 & 10 \end{array}\right)$$

(the first row, multiplied respectively by 2, 3 1, is subtracted from the following rows)

$$\sim \left(\begin{array}{cccc|c} 2 & -1 & -1 & 3 & 1 \\ 0 & 0 & 1 & -5 & 3 \\ 0 & 0 & 2 & -10 & 6 \\ 0 & 0 & 3 & -15 & 9 \end{array}\right)$$

(the second row, multiplied respectively by 2, 3, is subtracted from the following rows)

$$\left(\begin{array}{cccc|c} 2 & -1 & -1 & 3 & 1 \\ 0 & 0 & 1 & -5 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

The system has been reduced to an echelon system, which, after the rejection of two equations of the form $0 = 0$, reduces to the following:

$$\begin{aligned} 2x_1 - x_2 - x_3 + 3x_4 &= 1, \\ x_3 - 5x_4 &= 3. \end{aligned}$$

The unknowns x_1 and x_3 are principal, and x_2 and x_4 are free. The system is indeterminate. From the second equation we can find an expression for x_3 in terms of x_4; substituting it in the first equation, we find an expression for x_1 in terms of x_2 and x_4. The general solution has the form

$$\begin{aligned} x_1 &= \tfrac{1}{2}x_2 + x_4 + 2, \\ x_3 &= 5x_4 + 3. \end{aligned}$$

In the general solution, letting, for instance, $x_2 = 2$, $x_4 = -1$, we find one of the solutions:

$$x_1 = 2, \quad x_2 = 2, \quad x_3 - 2, \quad x_4 = -1,$$

which, to distinguish it from the general solution, is called a *partial solution* of the given system of equations.

EXAMPLE 4. To solve the system of equations

$$\begin{aligned} 2x_1 + 3x_2 + 5x_3 + x_4 &= 3, \\ 3x_1 + 4x_2 + 2x_3 + 3x_4 &= -2, \\ x_1 + 2x_2 + 8x_3 - x_4 &= 8, \\ 7x_1 + 9x_2 + x_3 + 8x_4 &= 0. \end{aligned}$$

We transform the matrix

$$\left(\begin{array}{cccc|c} 2 & 3 & 5 & 1 & 3 \\ 3 & 4 & 2 & 3 & -2 \\ 1 & 2 & 8 & -1 & 8 \\ 7 & 9 & 1 & 8 & 0 \end{array}\right)$$

(interchange the first and third rows, and the new first row, multiplied respectively by 3, 2, 7, is subtracted from the remaining rows)

$$\sim \left(\begin{array}{cccc|c} 1 & 2 & 8 & -1 & 8 \\ 0 & -2 & -22 & 6 & -26 \\ 0 & -1 & -11 & 3 & -13 \\ 0 & -5 & -55 & 16 & -56 \end{array}\right)$$

(divide the second row by -2 and, multiplying it respectively by 1 and 5, we add it to the following rows)

$$\sim \left(\begin{array}{cccc|c} 1 & 2 & 8 & -1 & 8 \\ 0 & 1 & 11 & -3 & 13 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 9 \end{array}\right)$$

(divide the fourth row by 9 and interchange it with the third row)

$$\sim \left(\begin{array}{cccc|c} 1 & 2 & 8 & -1 & 8 \\ 0 & 1 & 11 & -3 & 13 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

The system has been reduced to echelon form, and contains an equation of the form $0 = 1$, and so is inconsistent.

3. Cramer's rule.

If we are given a system of equations $a_{i1}x_1 + \ldots + a_{in}x_n = b_i$, $i = 1, \ldots, n$ (i.e. the number of equations is equal to the number of unknowns) and the determinant of this system $d = |a_{ij}|_1^n \neq 0$, then the system has a unique solution, which can be found by means of *Cramer's formula* $x_j = d_j/d$, $j = 1, \ldots, n$, where the determinant d_j is obtained from the determinant d by replacing the jth column by the column of free terms.

EXAMPLE 1. To solve the system of equations

$$x_1 - 2x_2 = 1,$$
$$3x_1 - 4x_2 = 7.$$

The determinant of the system is

$$d = \begin{vmatrix} 1 & -2 \\ 3 & -4 \end{vmatrix} = 2 \neq 0.$$

Therefore we can apply Cramer's rule. We find

$$d_1 = \begin{vmatrix} 1 & -2 \\ 7 & -4 \end{vmatrix} = 10, \qquad d_2 = \begin{vmatrix} 1 & 1 \\ 3 & 7 \end{vmatrix} = 4.$$

Hence

$$x_1 = \frac{d_1}{d} = 5, \qquad x_2 = \frac{d_2}{d} = 2.$$

In the case when the determinant d of the coefficients of the unknowns is equal to zero, Cramer's rule cannot be applied. From the general theory of linear equations (subsection 6) it follows that such a system will be either indeterminate or inconsistent.

EXAMPLE 2. To solve the system of equations

$$x_1 + x_2 + x_3 = 1,$$
$$x_1 + 2x_2 + 2x_3 = 2,$$
$$2x_1 + 3x_2 + 3x_3 = 3.$$

Here $d = d_1 = d_2 = d_3 = 0$, because the third row is equal to the sum of the first two. The system reduces to the echelon form

$$x_1 + x_2 + x_3 = 1,$$
$$x_2 + x_3 = 1,$$
$$0 = 0$$

and is indeterminate.

EXAMPLE 3. To solve the system of equations

$$x_1 + x_2 + x_3 = 1,$$
$$2x_1 + 2x_2 + 2x_3 = 1,$$
$$3x_1 + 3x_2 + 3x_3 = 2.$$

Here again $d = d_1 = d_2 = d_3 = 0$, because the third row is equal to the sum of the first two. The system reduces to the echelon form

$$x_1 + x_2 + x_3 = 1,$$
$$0 = 1,$$
$$0 = 0$$

and is inconsistent

If the determinant d of the coefficients of the unknowns is equal to zero, and at least one of the determinants occurring in the numerators of Cramer's formula is distinct from zero, then the system is inconsistent.

4. The rank of a matrix.

We consider a rectangular matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \dots a_{1n} \\ a_{21} & a_{22} \dots a_{2n} \\ \cdot & \cdot \cdot \cdot \cdot \cdot \\ a_{s1} & a_{s2} \dots a_{sn} \end{pmatrix}$$

of s rows and n columns (see Introduction).

The *row rank* (*column rank*) of the matrix **A** is the greatest number of linearly independent rows (columns) of **A**. (See section 1, subsection 3).

It turns out that the row rank of a matrix is equal to the column rank. This number is called *the rank of the matrix*. The rank of the transpose of **A** (p. 6) is the same as that of **A**.

An *elementary transformation* of the matrix **A** is a transformation of one of the following forms: (1) interchange of any two rows, (2) multiplication of a row by a number c distinct from zero, (3) addition to one row of another row multiplied by any number, and any analogous transformations of columns. Two matrices are said to be *equivalent* if it is possible to go from one to the other of them by means of a finite number of elementary transformations. The rank of a matrix is unaltered by elementary transformations; in other words, the ranks of equivalent matrices are equal.

An *echelon matrix* is a matrix which has the property that if in any of its rows the first element distinct from zero is in the kth position, then in all the following rows there are zeros in the first k positions, for instance:

$$\begin{pmatrix} 2 & 1 & 3 & 2 & 5 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 2 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Any matrix can be reduced to an echelon matrix by means of elementary transformations of its rows alone. The rank of an echelon matrix is equal to the number of its non-zero rows.

A *canonical matrix* is one in which all terms not of the principal diagonal are zero, all terms on the principal diagonal are zero or one, and all ones precede all zeros. For instance:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Any matrix can be reduced to canonical form by means of elementary transformations (of rows and columns). The rank of a canonical matrix is equal to the number of 1's in its principal diagonal.

The easiest method of calculating the rank of a matrix with numerical elements is to reduce it by elementary transformations to an echelon or canonical matrix.

EXAMPLE 1. To find the rank of the matrix

$$\begin{pmatrix} 2 & 3 & 5 & -3 & -2 \\ 3 & 4 & 3 & -1 & -3 \\ 5 & 6 & -1 & 3 & -5 \end{pmatrix}.$$

By means of elementary transformations we reduce this matrix to echelon form, joining successive matrices by the equivalence sign $\sim$. We subtract the first row from the second and interchange these rows:

$$\begin{pmatrix} 1 & 1 & -2 & 2 & -1 \\ 2 & 3 & 5 & -3 & -2 \\ 5 & 6 & -1 & 3 & -5 \end{pmatrix}$$

(from the second and third rows we subtract the first, multiplied respectively by 2 and 5)

$$\sim \begin{pmatrix} 1 & 1 & -2 & 2 & -1 \\ 0 & 1 & 9 & -7 & 0 \\ 0 & 1 & 9 & -7 & 0 \end{pmatrix}$$

(from the third row we subtract the second)

$$\sim \begin{pmatrix} 1 & 1 & -2 & 2 & -1 \\ 0 & 1 & 0 & -7 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The rank of the matrix is two, and it is easy to reduce the last matrix to canonical form. By subtracting the first column, multiplied by suitable numbers, from the following ones, we reduce to zero all the elements of the first row except the first one, while the elements of the remaining rows are unaltered. Then, by subtracting the second column, multiplied by suitable numbers, from the following ones, we reduce to zero all the elements of the second row, except the second one, and we obtain the canonical matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Another method of calculating the rank of a matrix is connected with the evaluation of determinants.

A theorem on the rank of a matrix. If a matrix has a minor of order r (see p. 31), distinct from zero, for which all the minors of order $r+1$ containing it (the bordering minors) are equal to zero, then the rank of this matrix is r.

From this it follows that the rank of a matrix (not all of whose elements are zero) is equal to the highest order of the non-zero minors of the matrix.

In calculating the rank of a matrix by means of the bordering method it is necessary to proceed from the lower orders to the higher. First, we seek a minor of the first order (i.e. an element of the matrix) or, at once, a minor of the second order distinct from zero. Then we calculate the minors of the following order which border it, until we find one of them which is distinct from zero. We proceed to calculate the minors bordering this last minor, and so on, until we find a minor of order r, distinct from zero, for which either all the bordering minors of order $r+1$ are equal to zero, or else such minors in general do not occur (if the matrix contains r rows or r columns). Then the rank of the matrix is equal to r.

EXAMPLE 2. To find the rank of the matrix

$$\begin{pmatrix} 1 & -2 & 3 & -2 & 4 \\ 2 & -4 & 5 & 1 & 7 \\ 1 & -2 & 1 & 8 & 2 \end{pmatrix}.$$

The minor of the first order in the upper left-hand corner is equal to $1 \neq 0$. The minor of the second order bordering it is equal to zero:

$$\begin{vmatrix} 1 & -2 \\ 2 & -4 \end{vmatrix} = 0.$$

Using the second row and the third column, we obtain the bordering minor

$$\begin{vmatrix} 1 & 3 \\ 2 & 5 \end{vmatrix} = -1 \neq 0.$$

We calculate the minors of the third order which border it. The second column cannot be adjoined, because it is proportional to the first, and we obtain a minor equal to zero. Adjoining the third row and either the fourth or fifth column, we obtain the minors

$$\begin{vmatrix} 1 & 3 & -2 \\ 2 & 1 & 1 \\ 1 & 1 & 8 \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} 1 & 3 & 4 \\ 2 & 5 & 7 \\ 1 & 1 & 2 \end{vmatrix},$$

which are equal to zero; this means that the rank of the matrix is two.

The rank of the sum and product of two matrices. The addition and multiplication of matrices is defined in Chapter II, section 1, subsection 1.

The rank of the sum of two (or any number) of matrices is not greater than the sum of their ranks ([20], Problem 626). Any matrix of rank r can be represented in the form

of the sum of r matrices of rank 1, but cannot be represented in the form of a sum of more than r such matrices ([20], Problem 627).

Any matrix $\mathbf{C}$ of rank r can be represented in the form of a product $\mathbf{C} = \mathbf{AB}$, where $\mathbf{A}$ consists of r linearly independent columns and $\mathbf{B}$ of r linearly independent rows ([20], Problem 939). If we denote the rank of the matrix $\mathbf{A}$ by r_A, then for the rank of the product $\mathbf{AB}$ of two square matrices $\mathbf{A}$ and $\mathbf{B}$ of order n we have:

Sylvester's inequality:

$$\begin{aligned}\text{Rank}\,(\mathbf{A}) + \text{Rank}\,(\mathbf{B}) - n &\leqslant \text{Rank}\,(\mathbf{AB}) \\ &\leqslant \min\big(\text{Rank}\,(\mathbf{A}),\ \text{Rank}\,(\mathbf{B})\big)\end{aligned}$$

([6], p. 61, or [20], Problem 931).

The same inequality remains true for the product of rectangular matrices, with the hypothesis that n denotes the number of columns of the matrix $\mathbf{A}$ and the number of rows of the matrix $\mathbf{B}$.

The rank of symmetric and skew-symmetric matrices. The definitions of these matrices are given on p. 109.

A *principal minor* of a matrix is any minor which stands at the intersection of rows and columns with the same numbers, i.e. which is situated symmetrically with respect to the principal diagonal of the matrix.

The rank of a symmetric or skew-symmetric matrix is equal to the highest order of non-zero principal minors. To be more precise, if we agree to regard a minor "of zero order" as equal to 1, then: if in a symmetric (or skew-symmetric) matrix there exists a principal minor D of order r, distinct from zero, and all the principal minors of order $r+1$ and $r+2$ which contain it (or simply of order $r+2$ in the case of a skew-symmetric matrix) are zero, then the rank of the matrix is r [3], pp. 59–60, or [20], Problems 631, 633).

If we denote by D_k the minor of the kth order which stands in the upper left-hand corner of the matrix (by definition $D_0 = 1$), then by making some permutation of the rows and the same permutation of the columns of a symmetric matrix we can obtain a symmetric matrix in which in the sequence of minors $D_0, D_1, D_2, \ldots, D_r$ there will not be two zeros one after the other and $D_r \neq 0$, $D_k = 0$ for all $k > r$.

In the case of a skew-symmetric matrix of rank r, in a similar way we can obtain the conditions: the minors D_0, $D_2, \ldots, D_r$ are distinct from zero, and all the remaining

minors D_k are equal to zero. The rank r of a skew-symmetric matrix is always an even number.

EXAMPLE 3. To find the rank of the matrix

$$\begin{pmatrix} 0 & 4 & 10 & 1 \\ 4 & 8 & 18 & 7 \\ 10 & 18 & 40 & 17 \\ 1 & 7 & 17 & 3 \end{pmatrix}.$$

Because the matrix is symmetric, it is sufficient simply to know the principal minors

$$D_0 = 1, \quad D_1 = 8,$$

$$D_2 = \begin{vmatrix} 0 & 4 \\ 4 & 8 \end{vmatrix} = -16, \quad D_3 = \begin{vmatrix} 0 & 4 & 10 \\ 4 & 8 & 18 \\ 10 & 18 & 40 \end{vmatrix} = 0.$$

The other principal minor containing D_2 will be

$$D_3' = \begin{vmatrix} 0 & 4 & 1 \\ 4 & 8 & 7 \\ 1 & 7 & 3 \end{vmatrix} = 0.$$

Finally,

$$D_4 = \begin{vmatrix} 0 & 4 & 10 & 1 \\ 4 & 8 & 18 & 7 \\ 10 & 18 & 40 & 17 \\ 1 & 7 & 17 & 3 \end{vmatrix}$$

(from the second and third rows we subtract the fourth, multiplied respectively by 4 and 10)

$$= \begin{vmatrix} 0 & 4 & 10 & 1 \\ 0 & -20 & -50 & -5 \\ 0 & -52 & -130 & -13 \\ 1 & 7 & 17 & 3 \end{vmatrix} = 0,$$

because there are proportional rows. This means that the rank of the matrix is equal to two.

5. The general theory of linear equations.

We indicate a method which enables us to formulate the properties of a system of linear equations in a general form without performing any preliminary transformations on the given system.

Suppose that we are given a system of linear equations

$$a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n = b_i, \quad i = 1, 2, \ldots, s. \qquad (1.4)$$

The matrix of the system of equation (1.4) is the matrix of the coefficients of the unknowns

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{s1} & a_{s2} & \dots & a_{sn} \end{pmatrix}.$$

The *augmented matrix* is the matrix obtained from the matrix of the system by adjoining the column of free terms, i.e. the matrix

$$\mathbf{B} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{s1} & a_{s2} & \dots & a_{sn} & b_s \end{pmatrix}.$$

The problem of whether the system (1.4) is consistent is solved by the following.

Kronecker–Capelli Criterion. In order that the system of linear equations (1.4) should be consistent, it is necessary and sufficient that the rank of the matrix **A** of the system should be equal to the rank of the augmented matrix **B**.

Let the rank of the matrix of the system be equal to $r \geqslant 1$. In this case, any minor of order r of the matrix of the system which is distinct from zero will be called a *principal minor of the system*, and any minor of order $r+1$, obtained from a principal minor by adjoining the column of corresponding free terms and one of the remaining rows (if it exists), will be called a *characteristic minor* for the given principal minor. The criterion of consistency can be formulated in other ways.

For a system of linear equations to be consistent, it is necessary and sufficient that there should exist a principal minor for which all the characteristic minors are equal to zero or else do not exist at all (the latter case holds when $r = s$).

A consistent system of linear equations has a unique solution if the rank r of the matrix of the system is equal to the number of unknowns n, and infinitely many solutions if $r < n$.

If here $r > 0$, then after the principal unknowns we can take any r unknowns, from whose coefficients in some r equations of the system we can construct a minor of order r, distinct from zero. The remaining $n - r$ unknowns are free,

and can be given any numerical values. With the given values of the free unknowns the values of the principal unknowns are defined uniquely. In other words, there exists a unique solution of a consistent system of linear equations when the free unknowns have given numerical values.

A rule for the solution of a system of linear equations.

(1) By calculating the rank of the matrix of the system and of the augmented matrix we decide the question of whether the system is consistent; if the system is consistent, then we find one of the principal minors $D \neq 0$ of order r, equal to the rank of both matrices.

(2) We take the r equations in which the minor D occurs; the remaining equations are rejected; the r unknowns whose coefficients occur in the principal minor D are regarded as principal and remain on the left, and the remaining $n - r$ unknowns are regarded as free and are carried over to the right-hand side.

(3) By the method of eliminating unknowns (subsection 2) or by Cramer's rule (subsection 3) we find expressions for the principal unknowns in terms of the free ones. The equations obtained are called the *general solution.*

(4) By assigning any numerical values to the free unknowns, we find from the general solution the corresponding values of the principal unknowns and at the same time we find a solution of the given system of equations, which, to distinguish it from the general solution, may be called a *particular solution.* In this way we can obtain any solution. For this the free unknowns in the general solution must be assigned the same values which they have in the given solution.

EXAMPLE. To find the general solution and one particular solution for the system of equations

$$\begin{aligned} 2x_1 - x_2 + x_3 + 2x_4 + 3x_5 &= 2, \\ 6x_1 - 3x_2 + 2x_3 + 4x_4 + 5x_5 &= 3, \\ 6x_1 - 3x_2 + 4x_3 + 8x_4 + 13x_5 &= 9, \\ 4x_1 - 2x_2 + x_3 + x_4 + 2x_5 &= 1. \end{aligned}$$

We write out the augmented matrix, separating the column of free terms by means of a line:

$$\left(\begin{array}{ccccc|c} 2 & -1 & 1 & 2 & 3 & 2 \\ 6 & -3 & 2 & 4 & 5 & 3 \\ 6 & -3 & 4 & 8 & 13 & 9 \\ 4 & -2 & 1 & 1 & 2 & 1 \end{array}\right).$$

We calculate the rank of the matrix of the system. The minor of the second order in its upper right-hand corner is

$$\begin{vmatrix} 2 & 3 \\ 4 & 5 \end{vmatrix} = -2 \neq 0.$$

A minor of order 3 containing it is

$$D = \begin{vmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 1 & 1 & 2 \end{vmatrix} = -1 \neq 0.$$

Adjoining to it the third row and the second column, we obtain a minor of the fourth order

$$\begin{vmatrix} -1 & 1 & 2 & 3 \\ -3 & 2 & 4 & 5 \\ -3 & 4 & 8 & 13 \\ -2 & 1 & 1 & 2 \end{vmatrix} = 0.$$

It is not necessary to adjoin the first column because it is proportional to the second column. This means that the rank of the matrix of the system is equal to three and D can be taken as a principal minor. To determine whether the system is consistent it is sufficient to calculate the unique characteristic minor

$$\begin{vmatrix} 1 & 2 & 3 & 2 \\ 2 & 4 & 5 & 3 \\ 4 & 8 & 13 & 9 \\ 1 & 1 & 2 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 & 2 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & -1 & -1 & -1 \end{vmatrix} = 0.$$

This means that the rank of the augmented matrix is also three and the system is consistent. We take x_3, x_4, x_5 as principal, and x_1, x_2 as free unknowns and we solve the system of equations

$$\begin{aligned} x_3 + 2x_4 + 3x_5 &= 2 - 2x_1 + x_2, \\ 2x_3 + 4x_4 + 5x_5 &= 3 - 6x_1 + 3x_2, \\ x_3 + x_4 + 2x_5 &= 1 - 4x_1 + 2x_2. \end{aligned}$$

by the method of eliminating unknowns. We find the general solution

$$\begin{aligned} x_3 &= -4x_1 + 2x_2 - 1, \\ x_4 &= 0, \\ x_5 &= 2x_1 - x_2 + 1. \end{aligned}$$

By letting, for instance, $x = 1, x = 2$, we find

$$x_3 = -1; \quad x_4 = 0; \quad x_5 = 1.$$

In order that the system of linear equations (1.4) should be consistent for fixed values of the coefficients a_{ij} of the unknowns and any values of the free terms b_i, it is necessary and sufficient that the rank r of the matrix **A** of the system should be equal to the number s of equations in the system. Here the system has a unique solution if the number of equations is equal to the number of unknowns, $s = n$, and infinitely

many solutions if $s < n$. In particular, in order that the system of linear equations (1.4) with fixed values of the coefficients a_{ij} of the unknowns should have a unique solution for any values of the free terms b_i, it is necessary and sufficient that the given system should be a *Cramer system*, i.e. that the number of equations should be equal to the number of unknowns and the determinant of the system should be distinct from zero.

6. Systems of homogeneous linear equations. Fundamental systems of solutions.

A linear equation is called *homogeneous* if its free term is equal to zero.

A system of homogeneous linear equations has the form

$$a_{i1}x_1 + \ldots + a_{in}x_n = 0, \quad i = 1, 2, \ldots, s. \qquad (1.5)$$

Every system of homogeneous linear equations has the *zero solution* $x_1 = x_2 = \ldots = x_n = 0$ and this means that it is consistent.

In order that a system of homogeneous linear equations should have a non-zero solution, it is necessary and sufficient that the rank r of this system (i.e. the rank of the matrix of the coefficients of unknowns) should be less than the number of unknowns n. From this it follows, in particular, that any system of homogeneous linear equations in which the number of equations is less than the number of unknowns, $s < n$, has a non-zero solution.

In order that a system of homogeneous linear equations in which the number of equations is equal to the number of unknowns, $s = n$, should have a non-zero solution, it is necessary and sufficient that its determinant should be zero.

Any solution

$$x_1 = c_1, \quad x_2 = c_2, \ldots, x_n = c_n$$

of a system of linear equations with n unknowns can be regarded as a row (or column) $(c_1, c_2, \ldots, c_n)$. Therefore such concepts as the sum of two solutions, the product of a solution by a number, linear combination of solutions, linear dependence or independence of a system of solutions (section 1, subsection 3) all have a meaning.

Properties of solutions of a system of homogeneous linear equations. (1) The sum of two solutions is a solution (2) the product of a solution by any number is a solution.

From these properties it follows that (3) a linear combination of solutions is a solution. In particular, if there exists at least one non-zero solution, then by multiplying it by arbitrary numbers we can obtain infinitely many solutions.

A *fundamental* (or *basic*) system of solutions for a system of homogeneous linear equations (1.5) is a linearly independent system of solutions, in terms of which any solution of the system (1.5) can be expressed linearly. If the rank r of the system of equations (1.5) is equal to the number n of unknowns then this system does not have a fundamental system of solutions, because its unique solution will be the zero solution, which forms a linearly dependent system. If $r < n$, then the system (1.5) has infinitely many fundamental systems of solutions, and each of them consists of $n - r$ solutions and any $n - r$ linearly independent solutions form a fundamental system.

A rule for constructing a fundamental system of solutions. We take any determinant M of order $n-r$, distinct from zero (for instance the determinant in which the elements of the principal diagonal are equal to 1, and the remaining elements are zero). The free unknowns are assigned in turn values equal to the elements of the first, second and so on rows of the determinant M, and each time we find from the general solution the corresponding values of the principal unknowns. The $n - r$ solutions obtained form a fundamental system. By changing the arbitrarily chosen determinant M, we can obtain all possible fundamental systems of solutions.

EXAMPLE 1. To find the general solution and a fundamental system of solutions for the system of equations

$$\begin{aligned} 3x_1 + 4x_2 + x_3 + 2x_4 + 3x_5 &= 0, \\ 5x_1 + 7x_2 + x_3 + 3x_4 + 4x_5 &= 0, \\ 4x_1 + 5x_2 + 2x_3 + x_4 + 5x_5 &= 0, \\ 7x_1 + 10x_2 + x_3 + 6x_4 + 5x_5 &= 0. \end{aligned}$$

The given system can be reduced to echelon form (subsection 2):

$$\begin{aligned} x_1 + x_2 + x_3 - x_4 + 2x_5 &= 0, \\ x_2 - 2x_3 + 5x_4 - 3x_5 &= 0, \\ x_4 &= 0. \end{aligned}$$

We regard x_1, x_2 and x_4 as the principal unknowns, and x_3 and x_5 as the free unknowns. From the second equation we find

$$x_2 = 2x + 3x_5.$$

Substituting this expression for x_2 in the first equation, we find:

$$x_1 = -3x_3 - 5x_5.$$

Thus the general solution has the form $x_1 = -3x_3 - 5x_5$, $x_2 = 2x_3 + 3x_5$, $x_4 = 0$. By giving the free unknowns in turn values equal to the elements of the rows of the determinant $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$, we find the fundamental system solutions:

x_1	x_2	x_3	x_4	x_5
-3	2	1	0	0
-5	3	0	0	1

If the rank of a system of homogeneous linear equations is one less than the number of unknowns: $r = n - 1$, then $n - r = 1$, i.e. a fundamental system consists of one solution and any non-zero solution forms a fundamental system. In this case any two solutions differ from each other only by a numerical factor.

As a solution of a system of $n - 1$ homogeneous linear equations with n unknowns, i.e. a system of the form

$$a_{i1}x_1 + \ldots + a_{in}x_n = 0, \quad i = 1, 2, \ldots, n-1,$$

we can take the set of all minors obtained from the matrix of the system by crossing out each column in turn, where these minors are taken with alternating signs. Thus, if

$$M_i = \begin{vmatrix} a_{11} & \ldots & a_{1,i-1} & a_{1,i+1} & \ldots & a_{1n} \\ a_{21} & \ldots & a_{2,i-1} & a_{2,i+1} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{n-1,1} & \ldots & a_{n-1,i-1} & a_{n-1,i+1} & \ldots & a_{n-1,n} \end{vmatrix}, \quad i = 1, 2, \ldots, n,$$

then

$$x_1 = M_1, \quad x_2 = -M_2, \quad x_3 = M_3, \quad \ldots, x_n = (-1)^{n+1}M_n.$$

Here, if not all these minors are equal to zero, then the solution obtained forms a fundamental system of solutions and all other solutions are proportional to it ([11], p. 83, or [20], Problem 748). Otherwise $r < n - 1$, and this method only gives the zero value.

EXAMPLE 2. To find the general solution and a fundamental system of solutions for the system of equations

$$3x - 2y + 4z = 0,$$
$$2x - 3y + 5z = 0.$$

We form the matrix of the system:

$$\begin{pmatrix} 3 & -2 & 4 \\ 2 & -3 & 5 \end{pmatrix}.$$

Therefore

$$x = \begin{vmatrix} -2 & 4 \\ -3 & 5 \end{vmatrix} = 2, \quad y = -\begin{vmatrix} 3 & 4 \\ 2 & 5 \end{vmatrix} = -7, \quad z = \begin{vmatrix} 3 & -2 \\ 2 & -3 \end{vmatrix} = -5.$$

Because this solution is non-zero, it forms a fundamental system of solutions. All the solutions are proportional to it. This means that the general solution has the form: $x = 2c, y = -7c, z = -5c$, where c is a parameter which can take any numerical value.

7. The connexion between the solutions of homogeneous and non-homogeneous systems.

If, in a non-homogeneous system of linear equations, we replace all the free terms by zero, then we obtain a homogeneous system, called the *reduced system* for the given non-homogeneous system. The solutions of the given non-homogeneous system and the reduced system corresponding to it are connected in the following way.

(1) The sum of any solution of the non-homogeneous system and any solution of its reduced system is a solution of the non-homogeneous system.

(2) The difference of any two solutions of the non-homogeneous system is a solution of its reduced system.

From these two properties it follows that:

(3) All the solutions of the non-homogeneous system can be obtained by adding to (any) one of its solutions all the solutions of its reduced system. In other words, the general solution of a non-homogeneous system can be obtained if to any partial solution of this system we add the general solution of its reduced system.

EXAMPLE. To find the general solution of the system of equations

$$\begin{aligned} 3x - 2y + 4z &= 5, \\ 2x - 3y + 5z &= 4. \end{aligned}$$

This system has the partial solution $x = y = z = 1$. The general solution of its reduced system was found in Example 2 of the preceding section and has the form: $x = 2c, y = -7c, z = -5c$, where c is a parameter which can take any values. Therefore the general solution of the given non-homogeneous system of equations has the form

$$x = 1 + 2c, \quad y = 1 - 7c, \quad z = 1 - 5c,$$

where c is an arbitrary parameter.

CHAPTER II

MATRICES AND QUADRATIC FORMS

1. Matrices

1. Operations with matrices.

On the rank of matrices see Chapter I, section 2, subsection 4. The definitions of a matrix, its rows and columns, a square matrix, its order, principal and secondary diagonal were given in the Introduction.

We will call a rectangular (in particular, a square) matrix of m rows and n columns an (m, n)-*matrix* or a matrix of *dimension* $m \times n$, and for brevity we denote it by $\mathbf{A} = (a_{ij})_{m,n}$, and a square matrix of order n by $A = (a_{ij})_1^n$. Two matrices with the same number of rows and the same number of columns will be called matrices of the same dimensions.

For matrices we can define three fundamental operations: addition, multiplication and the multiplication of a matrix by a number.

The *sum* of two matrices $\mathbf{A} = (a_{ij})_{m,n}$ and $\mathbf{B} = (b_{ij})_{m,n}$ of the same dimensions $m \times n$ is the matrix of the same dimensions

$$\mathbf{A} + \mathbf{B} = (a_{ij})_{m,n} + (b_{ij})_{m,n} = (a_{ij} + b_{ij})_{m,n}.$$

EXAMPLE 1.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 1 & -2 & 1 \\ 2 & 3 & -6 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 4 \\ 6 & 8 & 0 \end{pmatrix}.$$

The *product* of two matrices $\mathbf{A} = (a_{ij})_{m,n}$ and $\mathbf{B} = (b_{jk})_{n,p}$, which are given in a definite order (first and second) and whose dimensions are connected by the condition: the number of columns of the first matrix is equal to the number of rows of the second, is the matrix $\mathbf{C} = (c_{i,k})_{m,p}$, whose elements are determined by the equations

$$c_{ik} = \sum_{1=j}^{n} a_{ij} b_{jk}; \quad i = 1, 2, \ldots, m; \quad k = 1, 2, \ldots, p. \quad (2.1)$$

Thus, the element of the product of the matrices **A** and **B** which stands in the ith row and kth column is equal to the sum of the products of the elements of the ith row of the first matrix **A** by the corresponding elements of the kth column of the second matrix **B**. The product of the matrices **A** and **B** in the given order is denoted by **AB** or **A.B**. For square matrices of the same order the product **AB** (and also **BA**) is always defined.

EXAMPLE 2. If

$$\mathbf{A} = \begin{pmatrix} 4 & -1 \\ 5 & -2 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix},$$

then

$$\mathbf{AB} = \begin{pmatrix} 0 & 3 & 6 \\ -3 & 0 & 3 \end{pmatrix}.$$

The product **BA** is not defined.

The multiplication of matrices is closely connected with linear transformations of unknowns. Suppose that we are given two linear transformations of the unknowns:

$$x_i = \sum_{j=1}^{n} a_{ij} y_j, \qquad i = 1, 2, \ldots, m; \tag{2.2}$$

$$y_j = \sum_{k=1}^{p} b_{jk} z_k, \qquad j = 1, 2, \ldots, n. \tag{2.3}$$

Substituting the expression for y_j from the equations (2.3) in (2.2), we again obtain a linear transformation of the form

$$x_i = \sum_{k=1}^{p} c_{ik} z_k, \qquad i = 1, 2, \ldots, m. \tag{2.4}$$

It is called the result of the successive application or the *product* of the given transformations (2.2) and (2.3). If

$$\mathbf{A} = (a_{ij})_{m,n}, \qquad \mathbf{B} = (b_{jk})_{n,p}, \qquad \mathbf{C} = (c_{ik})_{m,p}$$

are the matrices of the coefficients of the transformations (2.2), (2.3) and their product respectively (2.4), then $\mathbf{AB} = \mathbf{C}$.

The *product of the matrix* $\mathbf{A} = (a_{ij})_{m,n}$ *by the number* c is the matrix $c\mathbf{A} = (ca_{ij})_{m,n}$, obtained by multiplying all the elements of the matrix **A** by the number.

The operations on matrices have the following properties (in all the following equations it is presupposed that all the operations occurring on one side of the equation have a meaning,

and from this it follows that the operations on the other side of the equation also have a meaning):

(1) Addition of matrices is commutative: $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$.

(2) Addition of matrices is associative: $\mathbf{A}+(\mathbf{B}+\mathbf{C})=(\mathbf{A}+\mathbf{B})+\mathbf{C}$.

From these two properties we can deduce that in any sum of a finite number of matrices the summands can be written in any order and the brackets indicating the order of the operations can be arranged arbitrarily. For instance, $[\mathbf{C}+(\mathbf{D}+\mathbf{A})]+\mathbf{B}=[(\mathbf{A}+\mathbf{B})+\mathbf{C}]+\mathbf{D}$, and can be written simply $\mathbf{A}+\mathbf{B}+\mathbf{C}+\mathbf{D}$.

(3) Addition has a unique inverse operation, *subtraction*. This means that for any two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same dimensions there exists a unique matrix $\mathbf{C}$ with the property $\mathbf{B}+\mathbf{C}=\mathbf{A}$. It is called the *difference* of the matrices $\mathbf{A}$ and $\mathbf{B}$ and is denoted by $\mathbf{A}-\mathbf{B}$.

Property (3) is equivalent to the following two:

(3a) There exists a unique matrix $\mathbf{0}$ of the given dimensions, called the *zero*, such that $\mathbf{A}+\mathbf{0}=\mathbf{A}$ for any matrix $\mathbf{A}$ of the same dimensions. In fact, the zero matrix is the matrix in which all the elements are equal to zero.

(3b) For any matrix $\mathbf{A}$ of the given dimensions there exists a unique matrix, $-\mathbf{A}$ of the same dimensions with the property $\mathbf{A}+(-\mathbf{A})=\mathbf{0}$. In fact, as the matrix $-\mathbf{A}$ we must take the matrix all of whose elements have the opposite sign to the corresponding elements of the matrix $\mathbf{A}$.

Multiplication of matrices is not commutative, i.e. does not have the transposing property. In the general case $\mathbf{AB}\neq\mathbf{BA}$ even for square matrices of the same order $n\geqslant 2$.

EXAMPLE 3.

$$\mathbf{A}=\begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix},\quad \mathbf{B}=\begin{pmatrix}2 & 1\\ 4 & 3\end{pmatrix},\quad \mathbf{AB}=\begin{pmatrix}10 & 7\\ 22 & 15\end{pmatrix},$$

$$\mathbf{BA}=\begin{pmatrix}5 & 8\\ 13 & 20\end{pmatrix},\quad \mathbf{AB}\neq\mathbf{BA}.$$

(4) Multiplication of matrices is associative: $\mathbf{A}(\mathbf{BC})=(\mathbf{AB})\mathbf{C}$.

(5) Multiplication of matrices is distributive with respect to addition:

$$(\mathbf{A}+\mathbf{B})\mathbf{C}=\mathbf{AC}+\mathbf{BC},\quad \mathbf{C}(\mathbf{A}+\mathbf{B})=\mathbf{CA}+\mathbf{CB}.$$

From this it follows that multiplication of matrices is distributive with respect to subtraction: $(\mathbf{A}-\mathbf{B})\mathbf{C}=\mathbf{AC}-\mathbf{BC}$, $\mathbf{C}(\mathbf{A}-\mathbf{B})=\mathbf{CA}-\mathbf{CB}$.

(6) $a(b\mathbf{A}) = (ab)\mathbf{A}$, where a, b are numbers and $\mathbf{A}$ is a matrix.

(7) $a(\mathbf{AB}) = (a\mathbf{A})\mathbf{B}$, where a is a number and $\mathbf{A}$, $\mathbf{B}$ are matrices. We shall denote the transpose of the matrix $\mathbf{A}$ by $\mathbf{A}'$. (See Chapter I, section 1, subsection 3.)

(8) For the transposition of the product of matrices we have the formula $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$, and an analogous formula is true for any finite number of factors.

We denote the determinant of a square matrix $\mathbf{A}$ by $|\mathbf{A}|$.

(9) The determinant of the product of two square matrices of the same order is equal to the product of their determinants:

$$|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|,$$

and similarly for any finite number of factors.

(10) *An expression for a minor of the product of two matrices in terms of the minors of the factors* (the notation was given on p. 31). If $\mathbf{A}$ is a matrix of dimensions $m \times n$, and $\mathbf{B}$ a matrix of dimensions $n \times p$, and $\mathbf{C} = \mathbf{AB}$, then

$$\mathbf{C}\begin{pmatrix} i_1 & i_2 & \dots & i_r \\ k_1 & k_2 & \dots & k_r \end{pmatrix} = \sum_{1 \leqslant j_1 < j_2 < \dots < j_r \leqslant n} \mathbf{A}\begin{pmatrix} i_1 & i_2 & \dots & i_r \\ j_1 & j_2 & \dots & j_r \end{pmatrix} \mathbf{B}\begin{pmatrix} j_1 & j_2 & \dots & j_r \\ k_1 & k_2 & \dots & k_r \end{pmatrix}$$

where $r \leqslant n$. All the minors of the matrix $\mathbf{C}$ of order $r > n$ are equal to zero ([6], p. 19).

When $m = p = r$ this equation reduces to the Binet–Cauchy theorem (p. 22).

EXAMPLE 4. Multiplying the non-zero column

$$\mathbf{A} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}$$

by the non-zero row $\mathbf{B} = (b_1\, b_2 \dots b_p)$, we obtain the matrix

$$\mathbf{C} = \mathbf{AB} = \begin{pmatrix} a_1b_1 & a_1b_2 & \dots & a_1b_p \\ a_2b_1 & a_2b_2 & \dots & a_2b_p \\ \dots & \dots & \dots & \dots \\ a_mb_1 & a_mb_2 & \dots & a_mb_p \end{pmatrix},$$

in which all the minors of orders higher than the first are equal to zero, but which has an element distinct from zero. Thus the rank of the matrix $\mathbf{C}$ is equal to one.

For operations with partitioned matrices see [6], Chapter II, section 5.

2. The identity matrix and the inverse matrix.

The *identity matrix* of order n is the square matrix $\mathbf{E}$ of order n which has the property $\mathbf{AE} = \mathbf{EA} = \mathbf{A}$ for any matrix $\mathbf{A}$ of order n. Thus the matrix $\mathbf{E}$ plays the same part in the multiplication of matrices as the number 1 plays in the multiplication of numbers. For any order n there exists a unique identity matrix of the given order. It has the form

$$\mathbf{E} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

A square matrix $\mathbf{A}$ is called *singular* (or *degenerate*) if its determinant is equal to zero, $|\mathbf{A}| = 0$, and *non-singular* (or *non-degenerate*) if $|\mathbf{A}| \neq 0$.

The *inverse matrix* for the square matrix $\mathbf{A}$ of order n is the square matrix $\mathbf{A}^{-1}$ of the same order which has the property $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{E}$, where $\mathbf{E}$ is the identity matrix of order n. A singular matrix does not have an inverse because from property (9) of the preceding section it follows that the product of a singular matrix by any square matrix of the same order will again be a singular matrix. Any non-singular matrix $\mathbf{A} = (a_{ij})_1^n$ has a unique inverse matrix, in fact:

$$\mathbf{A}^{-1} = \begin{pmatrix} \frac{A_{11}}{|\mathbf{A}|} & \frac{A_{21}}{|\mathbf{A}|} & \dots & \frac{A_{n1}}{|\mathbf{A}|} \\ \frac{A_{12}}{|\mathbf{A}|} & \frac{A_{22}}{|\mathbf{A}|} & \dots & \frac{A_{n2}}{|\mathbf{A}|} \\ \cdot & \cdot & \cdot & \cdot \\ \frac{A_{1n}}{|\mathbf{A}|} & \frac{A_{2n}}{|\mathbf{A}|} & \dots & \frac{A_{nn}}{|\mathbf{A}|} \end{pmatrix},$$

where A_{ij} is the cofactor of the element a_{ij} in the matrix $\mathbf{A}$ (p. 8).

The matrix which is the inverse of the matrix $\mathbf{A}^{-1}$ will be the matrix $\mathbf{A}$, i.e. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$. The determinants of two mutually inverse matrices are mutually inverse numbers: $|\mathbf{A}| \cdot |\mathbf{A}^{-1}| = 1$. The matrix which is the inverse of the product of two non-singular matrices is equal to the product of the matrices which are the inverses of the factors, taken in the inverse order:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1},$$

and similarly for any finite number of factors.

Methods of calculating an inverse matrix. (1) We use the form of the elements of the inverse matrix given above.

EXAMPLE 1. To find the inverse matrix for the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 4 & 5 \end{pmatrix}; \quad |\mathbf{A}| = -7.$$

We find

$$\mathbf{A}^{-1} = -\frac{1}{7}\begin{pmatrix} 5 & -3 \\ -4 & 1 \end{pmatrix} = \begin{pmatrix} -\frac{5}{7} & \frac{3}{7} \\ \frac{4}{7} & -\frac{1}{7} \end{pmatrix}.$$

This method is not convenient for matrices of a high order.

(2) We denote the elements of the ith column of the inverse matrix $\mathbf{A}^{-1}$ for the matrix $\mathbf{A} = (a_{ij})_1^n$ by $x_1, x_2, \ldots, x_n$. Equating the elements of the ith column in the matrix equation $\mathbf{A}\mathbf{A}^{-1} = \mathbf{E}$, we obtain the system of equations

$$a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n = \begin{cases} 0 \text{ when } k \neq i, \\ 1 \text{ when } k = i, \end{cases} \quad k = 1, 2, \ldots, n,$$

where there is a 1 on the right side of the ith equation, and in all the remaining equations there are zeros. Because $|\mathbf{A}| \neq 0$, then by Cramer's rule (Chapter I, section 2, subsection 3) this system has a unique solution, which can be found by the method of eliminating unknowns (Chapter I, section 2, subsection 2). The application of Cramer's formula here would reduce to the preceding method of calculating the inverse matrix.

EXAMPLE 2. To find the inverse matrix for the matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & \ldots & 1 \\ 1 & 0 & \ldots & 1 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & \ldots & 0 \end{pmatrix}.$$

For the elements of the ith column of the inverse matrix we have the system of equations

$$x_1 + \ldots + x_{k-1} + x_{k+1} + \ldots + x_n = \begin{cases} 0 \text{ when } k \neq i, \\ 1 \text{ when } k = i. \end{cases}$$

Adding all the equations and dividing by $n - 1$, we obtain

$$\sum_{k=1}^{n} x_k = \frac{1}{n-1}.$$

Subtracting from this each of the equations of the system, we find

$$x_1 = x_2 = \ldots = x_{i-1} = x_{i+1} = \ldots = x_n = \frac{1}{n-1};$$

$$x_i = \frac{1}{n-1} - 1 = \frac{2-n}{n-1}.$$

Whence

$$\mathbf{A}^{-1} = \frac{1}{n-1}\begin{pmatrix} 2-n & 1 & \ldots & 1 \\ 1 & 2-n & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & \ldots & 2-n \end{pmatrix}.$$

(3) Any non-singular matrix $\mathbf{A}$ can be reduced to the identity matrix $\mathbf{E}$ by means of elementary transformations (p. 40) of rows alone (or of columns alone). By applying the same sequence of transformations to the identity matrix $\mathbf{E}$, we obtain the inverse matrix $\mathbf{A}^{-1}$. It is convenient to perform the elementary transformations on $\mathbf{A}$ and $\mathbf{E}$ simultaneously, writing both matrices side by side, separated by a line ([20], Problem 933).

EXAMPLE 3. We will find by this method the inverse matrix for the matrix in Example 1 of this section. We join the matrices by means of the equivalence sign $\sim$:

$$\left(\begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 4 & 5 & 0 & 1 \end{array}\right)$$

(from the second row we subtract the first row multiplied by 4)

$$\sim \left(\begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 0 & -7 & -4 & 1 \end{array}\right)$$

(to the first row we add the second row multiplied by 3/7, and then we multiply the second row by $-1/7$)

$$\sim \left(\begin{array}{cc|cc} 1 & 0 & -\frac{5}{7} & \frac{3}{7} \\ 0 & 1 & \frac{4}{7} & \frac{1}{7} \end{array}\right); \quad \mathbf{A}^{-1} = \begin{pmatrix} -\frac{5}{7} & \frac{3}{7} \\ \frac{4}{7} & -\frac{1}{7} \end{pmatrix}.$$

The minors of the inverse matrix $\mathbf{B} = \mathbf{A}^{-1}$ of the non-singular matrix $\mathbf{A}$ of order n can be expressed in the following way in terms of the minors of the matrix $\mathbf{A}$. If

$$1 \leqslant \begin{matrix} i_1 < i_2 < \ldots < i_p \\ j_1 < j_2 < \ldots < j_p \end{matrix} \leqslant n,$$

then

$$\mathbf{B}\begin{pmatrix} i_1 & i_2 & \cdots & i_p \\ j_1 & j_2 & \cdots & j_p \end{pmatrix} = \frac{(-1)^{\sum_{s=1}^{p}(i_s+j_s)}\,\mathbf{A}\begin{pmatrix} j'_1 & j'_2 & \cdots & j'_{n-p} \\ i'_1 & i'_2 & \cdots & i'_{n-p} \end{pmatrix}}{|\mathbf{A}|},$$

where the system of indices $i'_1 < i'_2 < \ldots < i'_{n-p}$ supplements the system $i_1 < i_2 < \ldots < i_p$, and the system $j'_1 < j'_2 < \ldots < j'_{n-p}$ supplements the system $j_1 < j_2 < \ldots < j_p$ up to the complete system of indices $1, 2, \ldots, n$ ([6], p. 26 or [20], Problem 972).

There is a close connexion between the calculation of the inverse matrix and the solution of *matrix equations* of the form

$$\mathbf{AX} = \mathbf{B}, \quad \mathbf{YA} = \mathbf{B}, \tag{2.5}$$

where $\mathbf{A}$, $\mathbf{B}$ are given matrices and $\mathbf{X}$, $\mathbf{Y}$ are to be found. If $\mathbf{A}$ is a rectangular or non-singular square matrix, then the solution of these equations reduces to the solution of a system of linear equations for the elements of every column of the matrix $\mathbf{X}$ or every row of the matrix $\mathbf{Y}$; to obtain these equations it is necessary to equate the elements of the matrices on both sides of the equation. But if $\mathbf{A}$ is a non-singular matrix, then the solution is given by the formulae

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}, \quad \mathbf{Y} = \mathbf{B}\mathbf{A}^{-1}. \tag{2.6}$$

In particular, if the matrix $\mathbf{B}$ consists of one column (and also the unknown matrix $\mathbf{X}$ or $\mathbf{Y}$), then the first of the formulae (2.6) reduces to Cramer's formula (Chapter I, section 2, subsection 3). However, this formula is more convenient than Cramer's formula in the case when it is necessary to solve a series of systems of equations, which differ only on their right-hand sides. Then, having found the matrix $\mathbf{A}^{-1}$, it is easy to calculate $\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}$ for different special values of $\mathbf{B}$.

By using operations with matrices we can, under known conditions, reduce the calculation of a determinant of even order to the calculation of a single determinant of half its order. Let the determinant $\Delta = \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix}$ of order $2n$ be divided into the four cells $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ of order n. Then we have

Schur's formulae:

$$\Delta = |\mathbf{AD} - \mathbf{ACA}^{-1}\mathbf{B}|, \quad \text{if } |\mathbf{A}| \neq 0;$$
$$\Delta = |\mathbf{AD} - \mathbf{BD}^{-1}\mathbf{CD}|, \quad \text{if } |\mathbf{D}| \neq 0;$$
$$\Delta = |\mathbf{AD} - \mathbf{CB}|, \quad \text{if } \mathbf{AC} = \mathbf{CA};$$
$$\Delta = |\mathbf{AD} - \mathbf{BC}|, \quad \text{if } \mathbf{CD} = \mathbf{DC} \text{ ([6], p. 45).}$$

3. Powers of matrices. Polynomials in matrices. Commuting matrices.

A matrix $\mathbf{A}$ can be multiplied by itself if and only if it is square. All the matrices in this section (unless otherwise stated) are presumed to be square and of the same order n.

The powers of a matrix are defined thus: if k is a natural number greater than 1, then $\mathbf{A}^k$ is the product of k matrices equal to $\mathbf{A}$. From the associative rule for the multiplication of matrices (p. 54) it follows that this product has a unique meaning. Further, $\mathbf{A}^1 = \mathbf{A}$, $\mathbf{A}^0 = \mathbf{E}$, where $\mathbf{E}$ is the identity matrix. Powers with negative exponents are defined only for non-singular matrices. In fact, $\mathbf{A}^{-1}$ is the inverse matrix (p. 56), and for positive integral k, $k > 1$, the matrix $\mathbf{A}^{-k}$ is defined as either of the two equal matrices $(\mathbf{A}^k)^{-1} = (\mathbf{A}^{-1})^k$.

The definition of a root of a matrix and the concept, connected with it, of the power of a matrix with fractional exponent can be found in [6], Chapter 8, sections 6 and 7. Powers of a matrix with any real or complex exponent can be defined as special cases of the general concept of a *function of a matrix* ([6], Chapter 5).

For operations with powers of matrices the following rule is valid (where k and l are any integers; in the case of negative exponents the matrices are presupposed non-singular):

$$\mathbf{A}^k\mathbf{A}^l = \mathbf{A}^{k+l}, \tag{2.7}$$

$$(\mathbf{A}^k)^l = \mathbf{A}^{kl}, \tag{2.8}$$

$$(\mathbf{AB})^k = \mathbf{A}^k\mathbf{B}^k, \tag{2.9}$$

provided that

$$\mathbf{AB} = \mathbf{BA}.$$

If $\mathbf{AB} = \mathbf{BA}$, then Newton's binomial theorem is true

$$(\mathbf{A} + \mathbf{B})^m = \mathbf{A}^m + C_m^1\mathbf{A}^{m-1}\mathbf{B} + C_m^2\mathbf{A}^{m-2}\mathbf{B}^2 + \ldots + \mathbf{B}^m,$$

where m is a positive integer.

The matrices $\mathbf{A}$ and $\mathbf{B}$ are said to *commute* if both the products $\mathbf{AB}$ and $\mathbf{BA}$ have a meaning and $\mathbf{AB} = \mathbf{BA}$. Two commuting matrices are necessarily square and of the same order. If $\mathbf{A}$ commutes with $\mathbf{B}$, then $\mathbf{A}$ commutes with $\mathbf{B}^{-1}$, $\mathbf{A}^{-1}$ with $\mathbf{B}$ and $\mathbf{A}^{-1}$ with $\mathbf{B}^{-1}$, provided that the inverse matrices exist. From (2.7) it follows that all the powers of the same matrix $\mathbf{A}$ commute with each other.

A *scalar matrix* is a square matrix of the form $\mathbf{A} = c\mathbf{E}$, where $\mathbf{E}$ is the identity matrix. In order that the square matrix

A of order n should commute with all matrices of the same order, it is necessary and sufficient that it should be scalar ([11], p. 105). If the matrix **A** commutes with any diagonal matrix (see p. 72), the elements of whose principal diagonal are all different, then this matrix **A** is itself diagonal ([20], Problem 820).

The *value of the polynomial*

$$f(x) = c_0 x^m + c_1 x^{m-1} + \ldots + c_m$$

with numerical coefficients $c_0, c_1, \ldots, c_m$ in the matrix **A**, or the value of the polynomial $f(x)$ when $x = \mathbf{A}$, is the matrix

$$f(\mathbf{A}) = c_0 \mathbf{A}^m + c_1 \mathbf{A}^{m-1} + \ldots + c_m \mathbf{E},$$

where **E** is the identity matrix. Because the powers of the matrix **A** commute, it follows that all polynomials in the matrix **A** commute with each other and, in particular, commute with the matrix **A**.

The *quotient of the matrices* **A** and **B** has a meaning only for a non-singular divisor **B**, where this quotient has two values: a *left quotient* $\mathbf{B}^{-1}\mathbf{A}$ and a *right quotient* $\mathbf{A}\mathbf{B}^{-1}$. These two quotients coincide if and only if **A** and **B** commute. Only in this case can we use the notation for the quotient $\mathbf{A} \div \mathbf{B}$ or $\frac{\mathbf{A}}{\mathbf{B}}$.

A *rational function*

$$h(x) = \frac{f(x)}{g(x)}$$

(where $f(x)$ and $g(x)$ are polynomials with numerical coefficients) *of the matrix A* is the matrix

$$h(\mathbf{A}) = \frac{f(\mathbf{A})}{g(\mathbf{A})}.$$

This value $h(\mathbf{A})$ has a meaning if and only if the matrix $g(\mathbf{A})$ is non-singular. All rational functions of the matrix **A** commute with each other and with the matrix **A**.

The inverse matrix $\mathbf{A}^{-1}$ for the non-singular matrix **A**, and therefore also the rational function

$$h(\mathbf{A}) = \frac{f(\mathbf{A})}{g(\mathbf{A})} = f(\mathbf{A})[g(\mathbf{A})]^{-1}$$

can be represented in the form of a polynomial in the matrix **A** ([6], p. 98).

Under known conditions we can define the value of any function $f(x)$ of the matrix $\mathbf{A}$ as the value of some polynomial of this matrix ([6], Chapter 5, section 1). Therefore the value of any function of the matrix $\mathbf{A}$ is a matrix which commutes with the matrix $\mathbf{A}$.

Besides polynomials in the matrix $\mathbf{A}$, there may also exist other matrices which commute with $\mathbf{A}$. All the matrices which commute with a given matrix $\mathbf{A}$ are polynomials in $\mathbf{A}$ if and only if all the invariant factors (p. 67) of its characteristic matrix, except the last, are equal to unity or (which is the same thing) if the minors of the $(n-1)$th order of the characteristic matrix $(\lambda\mathbf{E}-\mathbf{A})$ are all mutually prime ([6], p. 184). Without this condition the matrices which commute with $\mathbf{A}$ are more than just the polynomials in the matrix $\mathbf{A}$, and in general more than the polynomials in any one matrix ([6], p. 184).

In the most general case the matrices $\mathbf{X}$ which commute with a given square matrix $\mathbf{A}$ of order n are found in the following way. Let $\mathbf{U}$ be any non-singular matrix which transforms the matrix $\mathbf{A}$ to the *Jordan matrix* $\mathbf{A}_0$, i.e. which satisfies the equation $\mathbf{A}=\mathbf{U}\mathbf{A}_0\mathbf{U}^{-1}$ (subsection 11). Then $\mathbf{X}=\mathbf{U}\mathbf{X}_0\mathbf{U}^{-1}$, where $\mathbf{X}_0$ is any matrix which commutes with the Jordan matrix $\mathbf{A}_0$. All matrices $\mathbf{X}_0$ can be found by means of the following rule: we call a rectangular (in particular, a square) matrix *properly upper triangular* if all its elements which lie below the diagonal from the upper left-hand corner to the lower right-hand corner are zero, and, as to the remaining elements, all the elements which lie along each line parallel to the given diagonal are equal to each other; for instance:

$$\begin{pmatrix} a & b & c \\ 0 & a & b \\ 0 & 0 & a \end{pmatrix}, \quad \begin{pmatrix} a & b & c \\ 0 & a & b \\ 0 & 0 & a \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & a & b & c \\ 0 & 0 & a & b \\ 0 & 0 & 0 & a \end{pmatrix}.$$

Suppose that the Jordan cells which stand in the principal diagonal of the matix $\mathbf{A}_0$, in the order from top to bottom, have orders $p_1, p_2, \ldots, p_s$ and that in their diagonals stand the numbers $\lambda_1, \lambda_2, \ldots, \lambda_s$ (some of which may be equal to each other). Drawing the horizontal and vertical lines which separate the Jordan cells, we divide the matrix $\mathbf{A}_0$ into square and rectangular cells. Let $\mathbf{X}_0=(\mathbf{X}_{\alpha\beta})_1^s$, be the corresponding division into cells of the matrix $\mathbf{X}_0$. Here $\mathbf{X}_{\alpha\beta}$ is a cell of dimensions $p_\alpha \times p_\beta$. The matrix $\mathbf{X}_0$ is defined by the conditions: $\mathbf{X}_{\alpha\beta}=0$ if $\lambda_\alpha \neq \lambda_\beta$, and $\mathbf{X}_{\alpha\beta}$ is some properly upper triangular

matrix if $\lambda_\alpha = \lambda_\beta$. The number of unknown parameters in the matrix $\mathbf{X}_0$ or the number of linearly independent matrices $\mathbf{X}$ which commute with $\mathbf{A}$ is defined by formula

$$N = n_1 + 3n_2 + 5n_3 + \ldots + (2t-1)n_t,$$

where $n_1 \geqslant n_2 \geqslant n_3 \geqslant \ldots \geqslant n_t < 0$ are the powers of the invariant factors of the matrix $\lambda\mathbf{E} - \mathbf{A}$, distinct from one ([6], pp. 183–4).

EXAMPLE. If the matrix $\lambda\mathbf{E} - \mathbf{A}$ has the invariant factors $E_1(\lambda) = \ldots = E_5(\lambda) = 1$, $E_6(\lambda) = (\lambda - \lambda_1)^2$, $E_7(\lambda) = (\lambda - \lambda_1)^3(\lambda - \lambda_2)^2$, then

$$\mathbf{X}_0 = \left(\begin{array}{ccc:ccc:cc} a & b & c & d & e & 0 & 0 \\ 0 & a & b & 0 & d & 0 & 0 \\ 0 & 0 & a & 0 & 0 & 0 & 0 \\ \hdashline 0 & f & g & h & k & 0 & 0 \\ 0 & 0 & f & 0 & h & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & l & m \\ 0 & 0 & 0 & 0 & 0 & 0 & l \end{array}\right).$$

The number of the parameters $a, b, \ldots, m$ is equal to $N = n_1 + 3n_2 = 5 + 3 \cdot 2 = 11$.

4. The connexion between the multiplication of matrices and elementary transformations. Decomposition of a matrix into the product of triangular matrices.

Every elementary transformation (p. 40) of the rows (or columns) of a rectangular matrix $\mathbf{A}$ of dimensions $m \times n$ can be replaced by a multiplication of the matrix $\mathbf{A}$, on the left (respectively, on the right) by a square non-singular matrix of order m (respectively, n), obtained from the identity matrix of the same order by the application of the same elementary transformation.

EXAMPLE.

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

By adding the first row multiplied by -2 to the second row, we obtain the matrix

$$\mathbf{B} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 0 \end{pmatrix}.$$

Performing the same transformation on the identity matrix of order 2 we obtain the matrix

$$\mathbf{P} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}.$$

Then the equation $\mathbf{PA} = \mathbf{B}$ must be satisfied, i.e.

$$\begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 0 \end{pmatrix}.$$

Any matrix $\mathbf{A}$ of dimensions $m \times n$ and rank r can be represented in the form $\mathbf{A} = \mathbf{PRQ}$, where $\mathbf{P}$ and $\mathbf{Q}$ are non-singular matrices respectively of orders m and n, and $\mathbf{R}$ is a matrix of dimensions $m \times n$, the first r elements of whose principal diagonal (starting from the upper left-hand corner) are equal to 1, and whose remaining elements are equal to zero.

An *upper* (*lower*) *triangular* matrix is a square matrix whose non-zero elements occur only in the principal diagonal and above (below) this diagonal. The sum, product, product by a number, and inverse matrix (in the non-singular case), of upper (lower) triangular matrices, will again be upper (lower) triangular matrices.

Any square matrix $\mathbf{A}$ can be represented in the form $\mathbf{A} = \mathbf{PR}$ or $\mathbf{A} = \mathbf{RQ}$, where $\mathbf{R}$, as desired, is an upper or a lower triangular matrix, and $\mathbf{P}$ and $\mathbf{Q}$ are non-degenerate matrices.

Any square matrix $\mathbf{A} = (a_{ij})_1^n$ of rank r, whose first r corner minors (i.e. minors which stand in the upper left-hand corner) are distinct from zero

$$d_k = A\begin{pmatrix} 1 & 2 & \dots & k \\ 1 & 2 & \dots & k \end{pmatrix} \neq 0; \quad k = 1, 2, \dots, r, \tag{2.10}$$

can be represented in the form of a product

$$A = BC, \tag{2.11}$$

where

$$B = \begin{pmatrix} b_{11} & 0 & \dots & 0 \\ b_{21} & b_{22} & \dots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix}$$

is a lower triangular matrix, and

$$C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ 0 & c_{22} & \dots & c_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & & c_{nn} \end{pmatrix}$$

is an upper triangular matrix. Here the first r elements of the principal diagonals of the matrices $\mathbf{B}$ and $\mathbf{C}$ can be taken as any which satisfy the conditions

$$\left.\begin{aligned} b_{11}\ c_{11} &= d_1, \\ b_{22}\ c_{22} &= \frac{d_2}{d_1}, \\ &\cdots\cdots \\ b_{rr}\ c_{rr} &= \frac{d_r}{d_{r-1}}. \end{aligned}\right\} \tag{2.12}$$

The choice of these elements uniquely defines all the elements of the first r columns of the matrix $\mathbf{B}$ and the first r rows of the matrix $\mathbf{C}$. These elements can be found by means of the formulae

$$\left.\begin{aligned} b_{ik} &= \frac{b_{kk}}{d_k} A\begin{pmatrix} 1 & 2 & \dots & k-1 & i \\ 1 & 2 & \dots & k-1 & k \end{pmatrix}, \\ c_{ki} &= \frac{c_{kk}}{d_k} A\begin{pmatrix} 1 & 2 & \dots & k-1 & k \\ 1 & 2 & \dots & k-1 & i \end{pmatrix}; \\ i &= k, k+1, \dots, n; \ k = 1, 2, \dots, r, \end{aligned}\right\} \tag{2.13}$$

or by means of the recurrence formulae

$$\left.\begin{aligned} b_{ik} &= \frac{a_{ik} - \sum_{j=1}^{k-1} b_{ij} c_{jk}}{c_{kk}}, \quad c_{ki} = \frac{a_{ki} - \sum_{j=1}^{k-1} b_{kj} c_{ji}}{b_{kk}}; \\ i &= k, k+1, \dots, n; \ k = 1, 2, \dots, r. \end{aligned}\right\} \tag{2.14}$$

In the case when $r < n$ all the elements of the last $n - r$ columns of the matrix $\mathbf{B}$ can be taken as equal to zero, and the elements of the last $n - r$ rows of the matrix $\mathbf{C}$ can be taken arbitrarily or else, on the contrary, the elements of the last $n - r$ columns of the matrix $\mathbf{B}$ can be taken arbitrarily, and the elements of the last $n - r$ rows of the matrix $\mathbf{C}$ can be taken as equal to zero ([6], p. 37–8).

If $\mathbf{A} = (a_{ij})_1^n$ is a symmetric matrix of rank r, whose corner minors satisfy the conditions (2.10), then $\mathbf{A}$ can be represented in the form

$$\mathbf{A} = \mathbf{B}\mathbf{B}', \tag{2.15}$$

where $\mathbf{B} = (b_{ij})_1^n$ is a lower triangular matrix and $\mathbf{B}'$ is the transpose of $\mathbf{B}$. The elements of the matrix $\mathbf{B}$ can be found by means of the formulae

$$\left.\begin{aligned} b_{ik} &= \frac{1}{\sqrt{(d_k d_{k-1})}} \mathbf{A}\begin{pmatrix} 1 & 2 & \dots & k-1 & i \\ 1 & 2 & \dots & k-1 & k \end{pmatrix}; \quad i = k, k+1, \dots, n, \\ & \qquad k = 1, 2, \dots, r; \\ b_{ik} &= 0; \quad k = r+1, r+2, \dots, n \end{aligned}\right\} \tag{2.16}$$

([6], p. 39).

5. Polynomial matrices, invariant factors and elementary divisors.

A *polynomial matrix* or a λ-*matrix* is a rectangular (in particular, a square) matrix $\mathbf{A}(\lambda) = (a_{ij}(\lambda))_{m,n}$, whose elements are polynomials in one variable λ with numerical coefficients. If all these coefficients belong to a given number field K (see Introduction), then we will call the matrix $\mathbf{A}(\lambda)$ a λ-*matrix over the field* K. *Elementary transformations* of the λ-matrix $\mathbf{A}(\lambda)$ are transformations of the following types:

(I) interchange of two rows,

(II) multiplication of a row by a number c, distinct from zero,

(III) addition to one row of another row, multiplied by any polynomial $f(\lambda)$, and the analogous transformations for columns.

Two λ-matrices $\mathbf{A}(\lambda)$ and $\mathbf{B}(\lambda)$ of the same dimensions $m \times n$ are called *equivalent*, in symbols $\mathbf{A}(\lambda) \sim \mathbf{B}(\lambda)$, if one can go from the matrix $\mathbf{A}(\lambda)$ to $\mathbf{B}(\lambda)$ by means of a series of a finite number of elementary transformations. The relation of equivalence possesses three fundamental properties:

(1) *reflexive*: every matrix is equivalent to itself, $\mathbf{A}(\lambda) \sim \mathbf{A}(\lambda)$

(2) *symmetric*: if $\mathbf{A}(\lambda) \sim \mathbf{B}(\lambda)$, then $\mathbf{B}(\lambda) \sim \mathbf{A}(\lambda)$;

(3) *transitive*: if $\mathbf{A}(\lambda) \sim \mathbf{B}(\lambda)$ and $\mathbf{B}(\lambda) \sim \mathbf{C}(\lambda)$, then $\mathbf{A}(\lambda) \sim \mathbf{C}(\lambda)$.

From this it follows that the totality of all λ-matrices of given dimensions $m \times n$ can be divided into non-intersecting classes of equivalent matrices, i.e. into classes such that any two matrices from one class are equivalent, and any two matrices from different classes are not equivalent to each other. There arises the problem of the canonical form of a λ-matrix, which is characteristic of a given class of equivalent λ-matrices.

Let p be the smaller of the numbers m, n. Then a *canonical* or *normal* λ-matrix of dimensions $m \times n$ is a matrix in which (1) the first p terms of the principal diagonal are polynomials $E_1(\lambda), \dots, E_p(\lambda)$ each of which has leading coefficient 1, (2) $E_i(\lambda)$ divides $E_{i+1}(\lambda)$ for $i = 1, 2, \dots, p-1$, (3) all other elements in the matrix are zero.

From these properties it follows that if among the polynomials $E_i(\lambda)$ there are polynomials of degree zero, then they are equal to 1 and stand at the beginning of the diagonal, if however there are any zeros, then they stand at the end of the principal diagonal. Thus the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & \lambda(\lambda-3) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

is canonical.

Every class of equivalent λ-matrices of dimensions $m \times n$ contains a unique canonical matrix, i.e. every λ-matrix $\mathbf{A}(\lambda)$ is equivalent to a unique canonical λ-matrix $\mathbf{C}(\lambda)$, called the *canonical form* or the *normal form* of the given matrix $\mathbf{A}(\lambda)$, and the polynomials $E_1(\lambda)$, $E_2(\lambda)$, ..., $E_p(\lambda)$, standing in the principal diagonal of $\mathbf{C}(\lambda)$, are called the *invariant factors* of the matrix $\mathbf{A}(\lambda)$. Here p is the smaller of the dimensions of the matrix $\mathbf{A}(\lambda)$ ([14], pp. 107–12).

The first method of calculating the invariant factors of a λ-matrix lies in reducing it to canonical form by means of elementary transformations.

EXAMPLE 1. To find the canonical form and invariant factors of the λ-matrix

$$\mathbf{A}(\lambda) = \begin{pmatrix} \lambda & \lambda-1 & \lambda-1 & \lambda-2 \\ \lambda & \lambda^3+\lambda & \lambda-1 & \lambda^3+\lambda-1 \\ \lambda+1 & \lambda^3+\lambda+2 & \lambda+1 & \lambda^3+2\lambda+3 \\ \lambda-1 & \lambda-3 & \lambda-3 & -6 \end{pmatrix}.$$

We subtract the third column from the remaining ones:

$$\mathbf{A}(\lambda) \sim \begin{pmatrix} 1 & 0 & \lambda-1 & -1 \\ 1 & \lambda^3+1 & \lambda-1 & \lambda^3 \\ 0 & \lambda^3+1 & \lambda+1 & \lambda^3+\lambda+2 \\ 2 & 0 & \lambda-3 & -\lambda-3 \end{pmatrix}$$

(from the second and fourth rows we subtract the first, multiplied respectively by 1 and 2)

$$\sim \begin{pmatrix} 1 & 0 & \lambda-1 & -1 \\ 0 & \lambda^3+1 & 0 & \lambda^3+1 \\ 0 & \lambda^3+1 & \lambda+1 & \lambda^3+\lambda+2 \\ 0 & 0 & -\lambda-1 & -\lambda-1 \end{pmatrix}$$

(from the third and fourth columns we subtract the first, multiplied respectively by $\lambda-1$ and -1)

$$\sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda^3+1 & 0 & \lambda^3+1 \\ 0 & \lambda^3+1 & \lambda+1 & \lambda^3+\lambda+2 \\ 0 & 0 & -\lambda-1 & -\lambda-1 \end{pmatrix}$$

(we subtract the second and third columns from the fourth)

$$\sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda^3+1 & 0 & 0 \\ 0 & \lambda^3+1 & \lambda+1 & 0 \\ 0 & 0 & -\lambda-1 & 0 \end{pmatrix}$$

(from the third row we subtract the second and add the row obtained to the fourth)

$$\sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda^3+1 & 0 & 0 \\ 0 & 0 & \lambda+1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

(we interchange the second and third rows, and also the second and third columns)

$$\sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda+1 & 0 & 0 \\ 0 & 0 & \lambda^3+1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

This is a canonical matrix; this means that the matrix $\mathbf{A}(\lambda)$ has the invariant factors:

$$E_1(\lambda) = 1, \quad E_2(\lambda) = \lambda + 1, \quad E_3(\lambda) = \lambda^3 + 1, \quad E_4(\lambda) = 0.$$

The polynomials in λ with coefficients in a numerical (or any) field K do not themselves form a field, but they are contained in the field of rational functions in λ with coefficients from K. Therefore the fundamental properties of rank, connected with the concepts of a minor and linear dependence, are true for λ-matrices. Thus, under elementary transformations of a λ-matrix its rank is unaltered and the theorem on rank of a matrix (p. 41) is true for λ-matrices.

Suppose that we are given a λ-matrix $\mathbf{A}(\lambda) = (a_{ij}(\lambda))_{m,n}$ of rank r and $p = \min(m, n)$. The *determinantal factors* of this matrix are the polynomials

$$D_1(\lambda), \quad D_2(\lambda), \quad \ldots, \quad D_p(\lambda),$$

where, when $k = 1, 2, \ldots, r$, the polynomial $\mathbf{D}_k(\lambda)$ is defined as the greatest common divisor (Chapter III) of all the minors of the kth order of the matrix $\mathbf{A}(\lambda)$, where the coefficient of its leading term is reduced to 1, and when $k = r+1, r+2, \ldots, p$, by definition, $\mathbf{D}_k(\lambda) = 0$. Under elementary transformations of the matrix $\mathbf{A}(\lambda)$ the determinantal factors are unaltered ([14], p. 109).

The second method of calculating the invariant factors of a λ*-matrix* is based on the following connexion between them and the determinantal factors:

$$D_k(\lambda) = E_1(\lambda)E_2(\lambda)\dots E_k(\lambda), \quad k = 1, 2, \dots, p. \tag{2.17}$$

From this we find

$$\left.\begin{aligned} &E_k(\lambda) = \frac{D_k(\lambda)}{D_{k-1}(\lambda)}, \quad k = 1, 2, \dots, r \\ &\text{(when } k = 1 \text{ we take it that } D_0(\lambda) = 1), \\ &E_k(\lambda) = 0, \quad k = r+1, r+2, \dots, p \end{aligned}\right\} \tag{2.18}$$

(see [14], p. 112). This method is convenient in cases when only a few elements of the λ-matrix depend on λ, as, for instance, in the characteristic matrix $\lambda\mathbf{E} - \mathbf{A}$.

EXAMPLE 2. To find the invariant factors and the canonical form of the λ-matrix

$$\mathbf{A}(\lambda) = \begin{pmatrix} \lambda - 3 & -1 & 1 \\ 0 & \lambda - 2 & 0 \\ -1 & -1 & \lambda - 1 \end{pmatrix}.$$

Because the matrix contains elements not depending on λ and distinct from zero, $\mathbf{D}_1(\lambda) = 1$. All the minors of order 2, containing the second row of the matrix, are divisible by $\lambda - 2$. We will calculate the remaining three minors.

$$M_{21} = \begin{vmatrix} -1 & 1 \\ -1 & \lambda - 1 \end{vmatrix} = -\lambda + 2, \quad M_{22} = \begin{vmatrix} \lambda - 3 & 1 \\ -1 & \lambda - 1 \end{vmatrix} = \lambda^2 - 4\lambda + 4,$$

$$M_{23} = \begin{vmatrix} \lambda - 3 & -1 \\ -1 & -1 \end{vmatrix} = -\lambda + 2.$$

Therefore $\mathbf{D}_2(\lambda) = \lambda - 2$. Expanding the determinant of the matrix $\mathbf{A}(\lambda)$ by the second row, we find

$$D_3(\lambda) = (\lambda - 2)M_{22} = (\lambda - 2)^3.$$

By the formulae (2.18) we find the invariant factors

$$E_1(\lambda) = 1, \quad E_2(\lambda) = \lambda - 2, \quad E_3(\lambda) = (\lambda - 2)^2.$$

The canonical form of the matrix $\mathbf{A}(\lambda)$ has the form

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda - 2 & 0 \\ 0 & 0 & (\lambda - 2)^2 \end{pmatrix}.$$

As in the case of numerical matrices (p. 63) every elementary transformation of rows (or columns) of the λ-matrix $\mathbf{A}(\lambda)$ of dimensions $m \times n$ can be replaced by a multiplication of this matrix on the left (respectively, on the right) by a so-called *elementary* λ-matrix, obtained by applying the given elementary

transformation to the identity matrix of order m (respectively, of order n). Elementary matrices are special cases of invertible λ-matrices.

The inverse matrix for a non-singular square λ-matrix in the general case will not be a λ-matrix, because it may have fractional elements.

A square λ-matrix is called *invertible* (or *unimodular*) if its inverse matrix is also a λ-matrix. In order that a square λ-matrix should be invertible, it is necessary and sufficient that its determinant should not contain λ and should be distinct from zero, for instance:

$$\begin{pmatrix} \lambda^2 & \lambda+2 \\ \lambda-2 & 1 \end{pmatrix}.$$

The product of several invertible λ-matrices is again invertible. The canonical form of an invertible λ-matrix is the identity matrix of the same order. Any invertible λ-matrix can be decomposed into the product of a finite number of elementary λ-matrices ([14], pp. 113–16).

A criterion for the equivalence of λ-matrices. In order that two λ-matrices $\mathbf{A}(\lambda)$ and $\mathbf{B}(\lambda)$ of the same dimensions should be equivalent, it is necessary and sufficient that one of the following conditions should be satisfied:

(1) they have the same set of determinantal factors;

(2) they have the same set of invariant factors;

(3) they are connected by the relation

$$\mathbf{B}(\lambda) = \mathbf{P}(\lambda)\,\mathbf{A}(\lambda)\mathbf{Q}(\lambda), \tag{2.19}$$

where $\mathbf{P}(\lambda)$ and $\mathbf{Q}(\lambda)$ are invertible λ-matrices ([14], p. 113).

An algorithm for calculating the matrices $\mathbf{P}(\lambda)$ and $\mathbf{Q}(\lambda)$. We reduce the λ-matrices $\mathbf{A}(\lambda)$ and $\mathbf{B}(\lambda)$ of the same dimensions $m \times n$ to a common canonical form $\mathbf{C}(\lambda)$ by means of elementary transformations (if the canonical forms of the given matrices turn out to be different, then these matrices are not equivalent). By performing in reverse order the transformations inverse to those which reduce $\mathbf{B}(\lambda)$ to $\mathbf{C}(\lambda)$, we obtain a chain of transformations reducing $\mathbf{C}(\lambda)$ to $\mathbf{B}(\lambda)$, and, by setting in front of this the chain of transformations which reduces $\mathbf{A}(\lambda)$ to $\mathbf{C}(\lambda)$, we obtain a chain of transformations which reduces $\mathbf{A}(\lambda)$ to $\mathbf{B}(\lambda)$. Such a chain of transformations can sometimes be constructed without using the canonical form $\mathbf{C}(\lambda)$. The matrix $\mathbf{P}(\lambda)$ is obtained from the identity

matrix $\mathbf{E}_m$ of order m by applying to it in the same order all the tranformations of the rows in the chain which reduces $\mathbf{A}(\lambda)$ to $\mathbf{B}(\lambda)$, and the matrix $\mathbf{Q}(\lambda)$ is obtained from the identity matrix $\mathbf{E}_n$ of order n by applying to it in the same order all the transformations of columns of the same chain. Neither the chain which reduces $\mathbf{A}(\lambda)$ to $\mathbf{B}(\lambda)$ nor the matrices $\mathbf{P}(\lambda)$ and $\mathbf{Q}(\lambda)$ are defined uniquely.

Elementary divisors. Let $\mathbf{A}(\lambda)$ be a λ-matrix over the number field K (see Introduction). The *elementary divisors* matrix $\mathbf{A}(\lambda)$ over the field K are the maximal powers of the factors irreducible over the field K (Chapter III) which occur in the decomposition of this matrix into invariant factors. Here, in the set of elementary divisors of the matrix $\mathbf{A}(\lambda)$, each of them occurs as many times as it occurs in the expansion of invariant factors.

Thus, if $E_{s+1}(\lambda)$, $E_{s+2}(\lambda)$, ..., $E_r(\lambda)$ are all the invariant factors of the matrix $\mathbf{A}(\lambda)$ having positive powers, and if their decompositions into powers of distinct factors irreducible over K have the form

$$\left.\begin{aligned} E_i(\lambda) = [p_{i1}(\lambda)]^{k_{i1}}[p_{i2}(\lambda)]^{k_{i2}} \dots [(p_{it_i}(\lambda)^{k_{it_i}}], \\ i = s+1,\ s+2, \dots,\ r, \end{aligned}\right\} \tag{2.20}$$

then the set of elementary divisors of this matrix consists of the polynomials

$$\begin{gathered} e_1(\lambda) = [p_{s+1,1}(\lambda)]^{k_{s+1,1}}, \\ e_2(\lambda) = [p_{s+1,2}(\lambda)]^{k_{s+1,2}}, \quad \dots, \quad e_t(\lambda) = [p_{r,t_r}(\lambda)]^{k_{rt_r}}, \end{gathered} \tag{2.21}$$

where $t = t_{s+1} + t_{s+2} + \dots + t_r$ is the number of elementary divisors.

If the field K is extended (see Chapter III) the determinantal factors $D_k(\lambda)$ and the invariant factors $E_k(\lambda)$ of the matrix $\mathbf{A}(\lambda)$ are unaltered, but the elementary divisors may change.

EXAMPLE 3. We will find the elementary divisors of the λ-matrix

$$\mathbf{A}(\lambda) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \lambda^2+2 & 0 & 0 & 0 \\ 0 & 0 & (\lambda^2-4)^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

over the fields of rational, real and complex numbers. The matrix $\mathbf{A}(\lambda)$ is in canonical form. Therefore, along its principal diagonal stand the invariant factors

$$E_1(\lambda) = 1, \quad E_2(\lambda) = \lambda^2 + 2, \quad E_3(\lambda) = (\lambda^4 - 4)^2, \quad E_4(\lambda) = 0.$$

Decomposing $E_2(\lambda)$ and $E_3(\lambda)$ into irreducible factors over each of the given fields, we obtain the system of elementary divisors:

(1) over the field of rational numbers:

$$e_1(\lambda) = \lambda^2 + 2, \quad e_2(\lambda) = (\lambda^2 + 2)^2, \quad e_3(\lambda) = (\lambda^2 - 2)^2;$$

(2) over the field of real numbers:

$$e_1(\lambda) = \lambda^2 + 2, \quad e_2(\lambda) = (\lambda^2 + 2)^2, \quad e_3(\lambda) = (\lambda + \sqrt{2})^2, \quad e_4(\lambda) = (\lambda - \sqrt{2})^2;$$

(3) over the field of complex numbers:

$$e_1(\lambda) = \lambda + 2i, \quad e_2(\lambda) = \lambda - 2i, \quad e_3(\lambda) = (\lambda + 2i)^2, \quad e_4(\lambda) = (\lambda - 2i)^2,$$

$$e_5(\lambda) = (\lambda + \sqrt{2})^2, \quad e_6(\lambda) = (\lambda - \sqrt{2})^2$$

Because over the field of complex numbers any polynomial of positive degree can be decomposed into linear factors (Chapter III), the elementary divisors of any λ-matrix over the field of complex numbers are powers of polynomials of the first degree, i.e. have the form $(\lambda - \lambda_0)^k$, where λ_0 is some number. Usually by elementary divisors (unless otherwise stated) we shall mean elementary divisors over the field of complex numbers.

The elementary divisors of the λ-matrix $\mathbf{A}(\lambda)$ are usually used to define its invariant factors of positive degree. The last of these invariant factors is equal to the product of the different irreducible factors with the highest degrees with which they occur in all the elementary divisors. The preceding invariant factor is formed by the same rule from the remaining elementary divisors, and so on. If the dimensions and rank of the matrix $\mathbf{A}(\lambda)$ are also given, they define all its invariant factors uniquely and this means also its canonical form.

EXAMPLE 4. To find the invariant factors of the λ-matrix $\mathbf{A}(\lambda)$ of dimensions 6×8 and rank 5, if we are given its elementary divisors $\lambda + 1$, $\lambda + 1$, $(\lambda + 1)^3$, $(\lambda - 1)^2$, $(\lambda - 1)^2$.

Because the number of invariant factors is equal to the smaller of the dimensions of the matrix, i.e. 6, and the number of the invariant factors distinct from zero is equal to the rank, i.e. 5, then $E_6(\lambda) = 0$. Further,

$$E_5(\lambda) = (\lambda + 1)^3(\lambda - 1)^2, \quad E_4(\lambda) = (\lambda + 1)(\lambda - 1)^2, \quad E_3(\lambda) = \lambda + 1.$$

Because we have already used all the elementary divisors, there cannot be any other invariant factors of positive degree. Therefore $E_2(\lambda) = E_1(\lambda) = 1$.

The method given above of calculating the invariant factors by means of the elementary divisors turns out to be useful in the case of diagonal and cellular diagonal matrices. A square matrix $\mathbf{A}$ is called *diagonal* if all its elements which do not lie along the principal diagonal are equal to zero. A diagonal matrix $\mathbf{A}$ is denoted by $\mathbf{A} = \{a_1, a_2, \ldots, a_n\}$ where the elements of the principal diagonal are given in the brackets.

Similarly, a square matrix $\mathbf{A}$ is called *cellular diagonal* if along its principal diagonal there stand the square matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_s$, and all the elements outside this chain of matrices are equal to zero. It is denoted by

$$\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_s\}.$$

The set of elementary divisors of a cellular diagonal (in particular, a diagonal) λ-matrix is the union of the sets of elementary divisors of its diagonal cells ([14], pp. 118–19). In this lies the advantage of elementary divisors over invariant factors.

EXAMPLE 5. To find the canonical form of the diagonal matrix $\mathbf{A}(\lambda) = \{\lambda^2, 0, \lambda^2 + \lambda, \lambda^2 - 1, \lambda^4 - \lambda^2\}$. Here the order of the matrix is equal to 5, and the rank is equal to 4. Therefore $E_5(\lambda) = 0$. The elementary divisors of the diagonal elements are:

$$\lambda^2,\ \lambda,\ \lambda + 1,\ \lambda + 1,\ \lambda - 1,\ \lambda^2,\ \lambda + 1,\ \lambda - 1.$$

This means that

$$E_4(\lambda) = \lambda^2(\lambda + 1)(\lambda - 1) = \lambda^4 - \lambda^2,\ E_3(\lambda) = \lambda^2(\lambda + 1)(\lambda - 1) = \lambda^4 - \lambda^2,$$
$$E_2(\lambda) = \lambda(\lambda + 1) = \lambda^2 + \lambda,\quad E_1(\lambda) = 1.$$

The canonical form is $\{1, \lambda^2 + \lambda, \lambda^4 - \lambda^2, \lambda^4 - \lambda^2, 0\}$.

Let $\mathbf{A}$ be a square matrix with elements from a given field K. The invariant factors and elementary divisors of its characteristic matrix $\lambda\mathbf{E} - \mathbf{A}$ (or $\mathbf{A} - \lambda\mathbf{E}$) are called the invariant factors and elementary divisors, respectively, of the matrix $\mathbf{A}$. All the invariant factors of the matrix $\mathbf{A}$ are distinct from zero, and the sum of their degrees is equal to the order n of $\mathbf{A}$. Therefore the sum of the degrees of the elementary divisors of the matrix $\mathbf{A}$ is also equal to its order. Because the rank of the matrix $\lambda\mathbf{E} - \mathbf{A}$ is equal to n, the invariant factors of the matrix $\mathbf{A}$ are uniquely determined by its elementary divisors. For the connexion between the elementary divisors of the matrix $\mathbf{A}$ and $f(\mathbf{A})$, where $f(\lambda)$ is a polynomial (or any function for which $f(\mathbf{A})$ has a meaning), see [6], Chapter 6, section 7.

6. Matrix polynomials.

A *matrix polynomial* is an expression of the form

$$\mathbf{F}(\lambda) = \mathbf{A}_0\lambda^s + \mathbf{A}_1\lambda^{s-1} + \ldots + \mathbf{A}_s,$$

where λ is a variable and $\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_s$ are square matrices with numerical elements all of the same order n. The number

n is called the *order of the polynomial* $\mathbf{F}(\lambda)$. If $\mathbf{A}_0 \neq 0$, then the number s is called the *degree of the matrix polynomial* $\mathbf{F}(\lambda)$. If the matrix $\mathbf{A}_0$ is not singular, i.e. $|\mathbf{A}_0| \neq 0$, then the matrix polynomial $\mathbf{F}(\lambda)$ is called *regular*. Any square λ-matrix can be represented in the form of a matrix polynomial, and conversely. For instance,

$$\begin{pmatrix} 2\lambda^2+\lambda+3 & \lambda-5 \\ 3\lambda+1 & \lambda^2-2 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\lambda^2 + \begin{pmatrix} 1 & 1 \\ 3 & 0 \end{pmatrix}\lambda + \begin{pmatrix} 3 & -5 \\ 1 & -2 \end{pmatrix}.$$

Two matrix polynomials of the same order can be added, subtracted and multiplied in the same way as polynomials with numerical coefficients, with the difference that the multiplication of numerical matrices, and therefore also matrix polynomials, is not necessarily commutative. The degree of the sum or difference of two matrix polynomials is not greater than the larger of their degrees. The degree of the product of two matrix polynomials is not greater than the sum of the degrees of the factors, and if at least one of the factors is regular, then it is equal to this sum.

Because the multiplication of matrix polynomials is non-commutative we can define two divisions with remainder—right and left. Suppose we are given two matrix polynomials

$$\mathbf{F}(\lambda) = \mathbf{A}_0\lambda^s + \mathbf{A}_1\lambda^{s-1} + \ldots + \mathbf{A}_s,$$

$$\mathbf{G}(\lambda) = \mathbf{B}_0\lambda^t + \mathbf{B}_1\lambda^{t-1} + \ldots + \mathbf{B}_t$$

of the same order n, where the polynomials $\mathbf{G}(\lambda)$ is regular. Then there exist a uniquely defined *right quotient* $\mathbf{P}(\lambda)$ and *right remainder* $\mathbf{R}(\lambda)$, satisfying the conditions

$$\left.\begin{array}{l} \mathbf{F}(\lambda) = \mathbf{P}(\lambda)\mathbf{G}(\lambda) + \mathbf{R}(\lambda), \\ \text{where either the degree of } \mathbf{R}(\lambda) \text{ is less than the degree of } \mathbf{G}(\lambda), \\ \qquad\qquad \text{or } \mathbf{R}(\lambda) = 0, \end{array}\right\} \quad (2.22)$$

and also a *left quotient* $\mathbf{Q}(\lambda)$ and *left remainder* $\mathbf{S}(\lambda)$, satisfying the conditions

$$\left.\begin{array}{l} \mathbf{F}(\lambda) = \mathbf{G}(\lambda)\mathbf{Q}(\lambda) + \mathbf{S}(\lambda), \\ \text{where either the degree of } \mathbf{S}(\lambda) \text{ is less than the degree of } \mathbf{G}(\lambda), \\ \qquad\qquad \text{or } \mathbf{S}(\lambda) = 0 \end{array}\right\} \quad (2.23)$$

([6], p. 70).

In right division the divisor stands on the right, and in left division on the left of the quotient.

Bezout's generalized theorem. In the division of a matrix polynomial $\mathbf{F}(\lambda)$ by the regular polynomial $\lambda\mathbf{E} - \mathbf{A}$ (or by $\mathbf{A} - \lambda\mathbf{E}$) the right remainder is equal to the *right value* of the dividend $\mathbf{F}(\lambda)$ when $\lambda = \mathbf{A}$, i.e. the matrix

$$\mathbf{F}^{(1)}(\mathbf{A}) = \mathbf{A}_0\mathbf{A}^s + \mathbf{A}_1\mathbf{A}^{s-1} + \ldots + \mathbf{A}_s,$$

and the left remainder is equal to the *left value* of the dividend $\mathbf{F}(\lambda)$ when $\lambda = \mathbf{A}$, i.e. the matrix $\mathbf{F}^{(2)}(\mathbf{A}) = \mathbf{A}^s\mathbf{A}_0 + \mathbf{A}^{s-1}\mathbf{A}_1 + \ldots + \mathbf{A}_s$.

From this it follows that $\mathbf{F}(\lambda)$ is divisible on the right (left) by $\lambda\mathbf{E} - \mathbf{A}$ or by $\mathbf{A} - \lambda\mathbf{E}$ if and only if $\mathbf{F}^{(1)}(\mathbf{A}) = 0$ (respectively $\mathbf{F}^{(2)}(\mathbf{A}) = 0$).

7. The characteristic roots and characteristic vectors of a matrix.†

The *characteristic matrix* of the square matrix $\mathbf{A} = (a_{ij})_1^n$ with numerical elements is the λ-matrix

$$\lambda\mathbf{E} - \mathbf{A} = \begin{pmatrix} \lambda - a_{11} & -a_{12} & \ldots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \ldots & -a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ -a_{n1} & -a_{n2} & \ldots & \lambda - a_{nn} \end{pmatrix},$$

where λ is an independent variable. The *characteristic polynomial* of the matrix $\mathbf{A}$ is the polynomial in λ which is equal to the determinant of the characteristic matrix, i.e.

$$|\lambda\mathbf{E} - \mathbf{A}| = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \ldots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \ldots & -a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ -a_{n1} & -a_{n2} & \ldots & \lambda - a_{nn} \end{vmatrix}.$$

The roots of the characteristic polynomial are called the *characteristic numbers* (or *characteristic roots* or *eigenvalues*) of the matrix $\mathbf{A}$. The set of all the characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_n$, where each number is written the same number of times as its multiplicity as a root of the characteristic poly-

† Here we give the concepts of the characteristic roots and the characteristic vectors of a matrix. The properties of characteristic vectors of linear transformations (or linear operators) can be found in any textbook of higher algebra (see, for instance, [6], [11]).

nomial, is called the *spectrum* of the matrix **A**. The equation $|\lambda \mathbf{E} - \mathbf{A}| = 0$ is called the *characteristic* (or *vector*) *equation* of the matrix **A**.

A *characteristic vector* of the matrix **A** belonging to the characteristic root λ_i is a non-zero vector $\mathbf{a} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ for which the column X, consisting of its coordinates, satisfies the matrix equation $\mathbf{AX} = \lambda_i \mathbf{X}$. This equation can be written in the form $(\lambda_i \mathbf{E} - \mathbf{A})\,\mathbf{X} = 0$. From this it is clear that the coordinates of the vector **a** are the solution of a homogeneous system of linear equations with matrix $\lambda_i \mathbf{E} - \mathbf{A}$, obtained from the characteristic matrix $\lambda \mathbf{E} - \mathbf{A}$ when $\lambda = \lambda_i$, i.e. the system

$$-a_{j1}x_1 - a_{j2}x_2 - \ldots + (\lambda_i - a_{jj})x_j - \ldots - a_{jn}x_n = 0,$$

$$j = 1, 2, \ldots, n.$$

In mathematical literature one often encounters another definition of the characteristic matrix—as the matrix $\mathbf{A} - \lambda \mathbf{E}$. In this case the characteristic polynomial is defined as $|\mathbf{A} - \lambda \mathbf{E}|$, the characteristic equation takes the form $|\mathbf{A} - \lambda \mathbf{E}| = 0$, and in the system of equations which is used for calculating the characteristic vectors, all the coefficients change sign. But all the essential properties of characteristic roots and characteristic vectors remain as before. In particualr, the spectrum of the matrix A and the set of characteristic vectors belonging to a given characteristic root λ_i remain unchanged.

The characteristic polynomial of the matrix **A** of order n is a polynomial of the nth degree in λ with leading coefficient equal to 1, and therefore is always distinct from zero. If we write it in the form

$$|\lambda E - A| = \lambda^n + c_1\lambda^{n-1} + \ldots + c_k\lambda^{n-k} + \ldots + c_n, \qquad (2.24)$$

then its coefficients can be expressed in terms of the elements of the matrix **A** in the following way:

$$c_k = (-1)^k S_k, \qquad (2.25)$$

where S_k is the sum of all the principal minors (i.e. the minors situated symmetrically with respect to the principal diagonal) of order k of the matrix **A**. In particular,

$$c_1 = -(a_{11} + a_{22} + \ldots + a_{nn}),\ c_n = (-1)^n |\mathbf{A}|. \qquad (2.26)$$

The sum of the diagonal elements $s_A = a_{11} + a_{22} + \ldots + a_{nn}$ is called the *trace* of the matrix **A**. Because by Vieta's formula

the roots of the characteristic polynomial (2.24) satisfy the equations

$$\lambda_1 + \lambda_2 + \ldots + \lambda_n = -c_1,$$

$$\lambda_1 \lambda_2 \ldots \lambda_n = (-1)^n c_n,$$

then from (2.26) it follows that

$$\lambda_1 + \lambda_2 + \ldots + \lambda_n = s_A, \quad \lambda_1 \lambda_2 \ldots \lambda_n = |\mathbf{A}|, \tag{2.27}$$

i.e. the sum of the characteristic roots of the matrix $\mathbf{A}$ is equal to its trace, and their product to its determinant. Similarly, from (2.25) it follows that, in general, the values of the elementary symmetric function σ_k (see Chapter III) of the characteristic roots of the matrix $\mathbf{A}$ is equal to the sum of the principal minors of order k of the matrix $\mathbf{A}$.

If by the characteristic polynomial we mean $|\mathbf{A} - \lambda\mathbf{E}|$, then the formulae (2.24)–(2.26) are altered thus:

$$|\mathbf{A} - \lambda\mathbf{E}| = (-1)^n \lambda^n + c_1 \lambda^{n-1} + \ldots + c_k \lambda^{n-k} + \ldots + c_n;$$

$$c_k = (-1)^{n-k} S_k; \quad c_1 = (-1)^{n-1}(a_{11} + a_{22} + \ldots + a_{nn}),$$

$$c_n = |\mathbf{A}|$$

([8], p. 108). The formulae (2.27) are unaltered.

From (2.27) it follows that the matrix $\mathbf{A}$ has at least one characteristic root equal to zero if and only if it is singular $|\mathbf{A}| = 0$. A square matrix $\mathbf{A}$ is called *nilpotent* if $\mathbf{A}^k = \mathbf{0}$ for some positive integral k. A matrix $\mathbf{A}$ is nilpotent if and only if all its characteristic roots are equal to zero.

The Hamilton–Cayley theorem. Any square numerical matrix $\mathbf{A}$ is a root of its own characteristic polynomial, i.e. annihilates it by the substitution $\lambda \to \mathbf{A}$: if $\varphi(\lambda) = |\lambda\mathbf{E} - \mathbf{A}|$, then $\varphi(\mathbf{A}) = 0$ ([14], p. 26, or [6], p. 74).

If $\lambda_1, \lambda_2, \ldots, \lambda_n$ is the spectrum of the matrix $\mathbf{A}$, then for a non-singular matrix $\mathbf{A}$ the spectrum of the inverse matrix $\mathbf{A}^{-1}$ consists of the numbers $\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_n^{-1}$ and, in general, for any integer k the spectrum of the matrix $\mathbf{A}^k$ has the form $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$; if, however, the matrix $\mathbf{A}$ is singular, then the last assertion remains true only for positive integers k. If $g(\lambda)$ is any polynomial with numerical coefficients and $\mathbf{A}$ is any square matrix of order n with spectrum $\lambda_1, \lambda_2, \ldots, \lambda_n$, then the matrix $g(\mathbf{A})$ has the spectrum $g(\lambda_1), g(\lambda_2), \ldots, g(\lambda_n)$ ([6], p. 75). Hence, by virtue of (2.27) we find that

$$|g(\mathbf{A})| = g(\lambda_1) g(\lambda_2) \ldots g(\lambda_n) = R(\varphi, g), \tag{2.28}$$

where $R(\varphi, g)$ is the resultant of the characteristic polynomial $\varphi(\lambda) = |\lambda\mathbf{E} - \mathbf{A}|$ and the given polynomial $g(\lambda)$ (Chapter III).

The *adjoint matrix* for the characteristic matrix $\lambda\mathbf{E} - \mathbf{A}$, where $\mathbf{A} = (a_{ij})_1^n$, is the matrix $\mathbf{B}(\lambda) = (b_{ij}(\lambda))_1^n$, where the element $b_{ij}(\lambda)$ is the cofactor of the element $a_{ji}(\lambda)$ of the jth row and ith column of the matrix $\lambda\mathbf{E} - \mathbf{A}$. The matrix $\mathbf{B}(\lambda)$ is a matrix polynomial in λ of degree $n - 1$. The equation

$$(\lambda\mathbf{E} - \mathbf{A})\mathbf{B}(\lambda) = \mathbf{B}(\lambda)(\lambda\mathbf{E} - \mathbf{A}) = |\lambda\mathbf{E} - \mathbf{A}|\mathbf{E}, \tag{2.29}$$

is true, where $\mathbf{E}$ is the identity matrix. Hence there is a connexion between the adjoint and inverse matrices:

$$\mathbf{B}(\lambda) = |\lambda\mathbf{E} - \mathbf{A}|(\lambda\mathbf{E} - \mathbf{A})^{-1}. \tag{2.30}$$

If we know the coefficients of the characteristic polynomial (2.24) we can calculate the adjoint matrix $\mathbf{B}(\lambda)$. Let $\varphi(\lambda) = |\lambda\mathbf{E} - \mathbf{A}|$. The quotient $h(\lambda, \mu) = \dfrac{\varphi(\lambda) - \varphi(\mu)}{\lambda - \mu}$ is a polynomial in the two variables λ and μ. Then $\mathbf{B}(\lambda) = h(\lambda\mathbf{E}, \mathbf{A})$ ([6], pp. 75–76).

If λ_0 is any characteristic root of the matrix $\mathbf{A}$ and $\mathbf{B}(\lambda_0) \neq 0$, then every non-zero column of the matrix $\mathbf{B}(\lambda_0)$ will be a characteristic vector of the matrix $\mathbf{A}$ belonging to the root λ_0.

D. K. Faddeyev has indicated a method of simultaneously calculating the coefficients of the characteristic polynomial and the adjoint matrix (see [6], p. 77).

8. The minimal polynomial of a matrix.

An *annihilator polynomial* for a square matrix $\mathbf{A}$ of order n is a polynomial $f(\lambda)$ with numerical coefficients, for which $f(\mathbf{A}) = 0$. The zero polynomial is an annihilator for any matrix $\mathbf{A}$. Any matrix $\mathbf{A}$ of order n has an annihilator polynomial of the nth degree; by the Hamilton–Cayley theorem (p. 77) such a polynomial is the characteristic polynomial of the matrix $\mathbf{A}$, i.e. $\varphi(\lambda) = |\lambda\mathbf{E} - \mathbf{A}|$. Any matrix $\mathbf{A}$ has infinitely many annihilator polynomials. The *minimal polynomial* of a square matrix $\mathbf{A}$ of order n is the non-zero annihilator polynomial for the matrix $\mathbf{A}$ of minimal degree, taken with leading coefficient equal to 1. For any matrix $\mathbf{A}$ of order n there exists a minimal polynomial, which is defined uniquely and has degree not greater than n.

EXAMPLE 1. The minimal polynomial of the zero matrix is equal to 1, and that of the identity matrix $\mathbf{E}$ is equal to $\lambda - 1$.

A polynomial $f(\lambda)$ is an annihilator for the matrix $\mathbf{A}$ if and only if $f(\lambda)$ is divisible by the minimal polynomial $\psi(\lambda)$ of this matrix. In particular, the characteristic polynomial $\varphi(\lambda)$ of the matrix $\mathbf{A}$ is always divisible by the minimal polynomial $\psi(\lambda)$ of this matrix. Some power of the minimal polynomial, for instance $[\psi(\lambda)]^n$, where n is the order of the matrix $\mathbf{A}$, is divisible by the characteristic polynomial $\varphi(\lambda)$ of the same matrix.

EXAMPLE 2. To find the minimal polynomial of an involutory matrix $\mathbf{A}$, i.e. a matrix for which $\mathbf{A}^2 = \mathbf{E}$.

In the given case $\mathbf{A}^2 - \mathbf{E} = 0$. Therefore an annihilator polynomial will be $f(\lambda) = \lambda^2 - 1$. Because the minimal polynomial divides any annihilator, the possible cases are:

$$\psi(\lambda) = \lambda - 1, \quad \psi(\lambda) = \lambda + 1, \quad \psi(\lambda) = \lambda^2 - 1.$$

In the first case $\mathbf{A} - \mathbf{E} = \mathbf{0}$, $\mathbf{A} = \mathbf{E}$. In the second case $\mathbf{A} + \mathbf{E} = \mathbf{0}$, $\mathbf{A} = -\mathbf{E}$. In the third case $\mathbf{A} \neq \pm \mathbf{E}$. Thus, if $\mathbf{A} \neq \pm \mathbf{E}$ and $\mathbf{A}^2 = \mathbf{E}$, then $\psi(\lambda) = \lambda^2 - 1$.

If $\mathbf{D}_{n-1}(\lambda)$ is the greatest common divisor of the minors of order $n-1$ of the characteristic matrix $\lambda\mathbf{E} - \mathbf{A}$, and $\mathbf{D}_n(\lambda) = |\lambda\mathbf{E} - \mathbf{A}|$, then the minimal polynomial $\psi(\lambda)$ of the matrix $\mathbf{A}$ can be found by means of the formula

$$\psi(\lambda) = \frac{\mathbf{D}_n(\lambda)}{\mathbf{D}_{n-1}(\lambda)} = \mathbf{E}_n(\lambda), \tag{2.31}$$

where $\mathbf{E}_n(\lambda)$ is the last invariant factor of the matrix $\lambda\mathbf{E} - \mathbf{A}$ (p. 67) ([6], p. 81).

From this it is clear that the minimal polynomial coincides with the characteristic polynomial if and only if $\mathbf{D}_{n-1}(\lambda) = 1$, i.e. if the matrix $\lambda\mathbf{E} - \mathbf{A}$ has only one invariant factor distinct from 1, namely $\mathbf{E}_n(\lambda)$.

The minimal polynomial of the cellular diagonal matrix

$$\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_s\}$$

is equal to the least common multiple of the minimal polynomials of the diagonal cells $\mathbf{A}_i$.

EXAMPLE 3. To find the minimal polynomial of the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

We have to find the minimal polynomials of the diagonal cells. For $\mathbf{A}_1 = (1)$

$$\psi_1(\lambda) = \lambda - 1,$$

for $\mathbf{A}_2 = (2)$

$$\psi_2(\lambda) = \lambda - 2,$$

for $\mathbf{A}_3 = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$ we look for the determinantal factor: $\mathbf{D}_1(\lambda) = 1$, $\mathbf{D}_2(\lambda) = (\lambda - 2)^2$. By the formula (2.31)

$$\psi_3(\lambda) = \frac{\mathbf{D}_2(\lambda)}{\mathbf{D}_1(\lambda)} = (\lambda - 2)^2.$$

The minimal polynomial of the matrix $\mathbf{A}$ is equal to the least common multiple of the polynomials $\psi_1(\lambda)$, $\psi_2(\lambda)$, $\psi_3(\lambda)$, i.e. $\psi(\lambda) = (\lambda - 1)(\lambda - 2)^2$.

The reduced adjoint matrix. Let $\mathbf{D}_{n-1}(\lambda)$ be the greatest common divisor of the minors of order $n - 1$ of the matrix $\lambda\mathbf{E} - \mathbf{A}$. Because all the elements of the adjoint matrix $\mathbf{B}(\lambda)$ (p. 78) differ from these minors only, if at all, in sign, then $\mathbf{D}_{n-1}\lambda$ will be the greatest common divisor of all the elements of the matrix $\mathbf{B}(\lambda)$. Therefore

$$\mathbf{B}(\lambda) = \mathbf{D}_{n-1}(\lambda)\mathbf{C}(\lambda), \tag{2.32}$$

where $\mathbf{C}(\lambda)$ is a λ-matrix and is called the *reduced adjoint matrix* for the characteristic matrix $\lambda\mathbf{E} - \mathbf{A}$.

From equations (2.31), (2.32) and (2.29) we find

$$(\lambda\mathbf{E} - \mathbf{A})\mathbf{C}(\lambda) = \mathbf{C}(\lambda)(\lambda\mathbf{E} - \mathbf{A}) = \psi(\lambda)\mathbf{E}. \tag{2.33}$$

Hence, connecting the matrix $\mathbf{C}(\lambda)$ with the inverse matrix:

$$\mathbf{C}(\lambda) = \psi(\lambda)(\lambda\mathbf{E} - \mathbf{A})^{-1}. \tag{2.34}$$

If we know the minimal polynomial $\psi(\lambda)$ of the matrix $\mathbf{A}$, we can find the reduced adjoint matrix $\mathbf{C}(\lambda)$ of the matrix $\lambda\mathbf{E} - \mathbf{A}$. In fact, let $g(\lambda, \mu) = \dfrac{[\psi(\lambda) - \psi(\mu)]}{(\lambda - \mu)}$ be a polynomial in the two variables λ, μ. Then

$$\mathbf{C}(\lambda) = g(\lambda\mathbf{E}, \mathbf{A}) \tag{2.35}$$

([6], p. 81).

Going over to determinants in equation (2.33), we obtain

$$|\lambda\mathbf{E} - \mathbf{A}|\,|\mathbf{C}(\lambda)| = [\psi(\lambda)]^n. \tag{2.36}$$

If λ_0 is any root of the characteristic polynomial of the matrix $\mathbf{A}$, then (as distinct from $\mathbf{B}(\lambda)$), $\mathbf{C}(\lambda_0) \neq 0$ and any non-zero column of the matrix $\mathbf{C}(\lambda_0)$ will be a characteristic vector of the matrix $\mathbf{A}$, belonging to the root λ_0 ([6], p. 81).

9. Similar matrices.

A square matrix **A** is said to be *similar* to the matrix **B** if there exists a non-singular matrix **T** for which

$$\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}. \tag{2.37}$$

Here, if the elements of the matrices **A** and **B** belong to a number field K (see Introduction), then the transforming matrix **T** can always also be taken with elements from the field K. One also says that the matrix **A** is *transformed* into the matrix **B** by means of the matrix **T**.

The relation of similarity possesses the three fundamental properties: reflexive: **A** is similar to **A**; symmetric: if **A** is similar to **B**, then **B** is also similar to **A**; transitive: if **A** is similar to **B** and **B** is similar to **C**, then **A** is similar to **C**. By virtue of the symmetry we can speak of the similarity of two matrices without specifying their order. From these properties it follows that the set of all matrices of a given order with elements from a field K can be divided into classes of similar matrices, and no matrices from different classes can be similar to each other. There arises the problem of a canonical form of the matrix which characterizes a given class of similar matrices. This problem is considered below in subsections 10–12.

A criterion for the similarity of matrices. In order that two matrices **A** and **B** should be similar, it is necessary and sufficient that their characteristic matrices $\lambda\mathbf{E} - \mathbf{A}$ and $\lambda\mathbf{E} - \mathbf{B}$ (regarded as λ-matrices) should be equivalent (p. 82) ([14], p. 123, or [6], p. 124).

By virtue of the criterion of the equivalence of λ-matrices (p. 70) and the properties of elementary divisors of characteristic matrices (p. 73), for the matrices **A** and **B** (of the same order) to be similar each of the following conditions is necessary and sufficient:

(1) matrices **A** and **B** have identical invariant factors;

(2) matrices **A** and **B** have identical elementary divisors;

(3) there exist invertible λ-matrices $\mathbf{P}(\lambda)$ and $\mathbf{Q}(\lambda)$, satisfying the equation

$$\lambda\mathbf{E} - \mathbf{B} = \mathbf{P}(\lambda)\,(\lambda\mathbf{E} - \mathbf{A})\,\mathbf{Q}(\lambda). \tag{2.38}$$

Thus, in order to discover whether the matrices **A** and **B** are similar, it is sufficient to find the invariant factors of the matrices $\lambda\mathbf{E} - \mathbf{A}$ and $\lambda\mathbf{E} - \mathbf{B}$ by any of the methods given in subsection 5.

It is more difficult to find the transforming matrix **T** satisfying the equation (2.37). There are infinitely many such

matrices; if $\mathbf{T}_0$ is any such matrix, then all the transforming matrices can be found from the equation $\mathbf{T} = \mathbf{C}\mathbf{T}_0$, where $\mathbf{C}$ is any non-singular matrix which commutes with $\mathbf{A}$. The method for finding all the matrices which commute with $\mathbf{A}$ is given on p. 62.

First method for constructing a transforming matrix. From (2.37) we find that $\mathbf{AT} = \mathbf{TB}$. Equating the elements, we obtain n^2 homogeneous linear equations for the elements t_{ij} of the matrix $\mathbf{T}$. We can take any solution of this system for which the matrix $\mathbf{T}$ is non-singular. To do this we can first find the general solution of the system, and replace all the principal unknowns in the determinant $|t_{ij}|_1^n$ by their expressions in terms of the free unknowns; we obtain a polynomial which is not identically zero. Therefore we can find values of the free unknowns for which this polynomial is distinct from zero. These values define all the elements of the matrix $\mathbf{T}$ uniquely.

EXAMPLE. To prove that the matrices

$$\mathbf{A} = \begin{pmatrix} 1 & -3 \\ 1 & 2 \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} 4 & -3 \\ 3 & -1 \end{pmatrix}$$

are similar, and to find the transforming matrix $\mathbf{T}$ for which $\mathbf{B} = \mathbf{T}^{-1}\mathbf{AT}$. The invariant factors of the characteristic matrices $\lambda\mathbf{E} - \mathbf{A}$ and $\lambda\mathbf{E} - \mathbf{B}$ are identical:

$$\mathbf{E}_1(\lambda) = 1, \ \mathbf{E}_2(\lambda) = \lambda^2 - 3\lambda + 5,$$

therefore the matrices $\mathbf{A}$ and $\mathbf{B}$ are similar. Setting

$$\mathbf{T} = \begin{pmatrix} x & y \\ z & t \end{pmatrix}$$

and equating the elements of the matrices in the equation $\mathbf{AT} = \mathbf{TB}$, we obtain

$$\begin{aligned} x - 3z &= 4x + 3y, \\ x + 2z &= 4z + 3t, \\ y - 3t &= -3x - y, \\ y + 2t &= -3z - t. \end{aligned}$$

Solving these by the method of eliminating unknowns, we find two linearly independent equations:

$$\begin{aligned} x + y + z &= 0, \\ y + 3z + 3t &= 0. \end{aligned}$$

The general solution is: $x = 2z + 3t$, $y = -3z - 3t$. Substituting these values for x and y in the determinant, we find

$$|T| = \begin{vmatrix} 2z + 3t & -3z - 3t \\ z & t \end{vmatrix} = 3z^2 + 5zt + 3t^2.$$

Taking, for instance, $z=-1$, $t=1$, we obtain $|\mathbf{T}|=1$, $x=1$, $y=0$ whence

$$\mathbf{T}=\begin{pmatrix}1 & 0\\ -1 & 1\end{pmatrix}.$$

Second method for constructing the transforming matrix. If the matrices $\mathbf{A}$ and $\mathbf{B}$ are similar, then their characteristic matrices $\lambda\mathbf{E}-\mathbf{A}$ and $\lambda\mathbf{E}-\mathbf{B}$ are equivalent. Therefore there exist invertible λ-matrices $\mathbf{P}(\lambda)$ and $\mathbf{Q}(\lambda)$, satisfying equation (2.38). The method for constructing them was given on p. 70. In the given case it is sufficient to find simply the matrix $\mathbf{Q}(\lambda)$. As the desired matrix $\mathbf{T}$ we can take the remainder after the division of the matrix $\mathbf{Q}(\lambda)$ on the right by the matrix $\lambda\mathbf{E}-\mathbf{B}$ (p. 74). If $\mathbf{Q}(\lambda)=\mathbf{C}_0\lambda^s+\mathbf{C}_1\lambda^{s-1}+\ldots+\mathbf{C}_s$, then by Bezout's generalized theorem (p. 75)

$$\mathbf{T}=\mathbf{Q}^{(1)}(\mathbf{B})=\mathbf{C}_0\mathbf{B}^s+\mathbf{C}_1\mathbf{B}^{s-1}+\ldots+\mathbf{C}_s.$$

However, in practice it is more convenient to find the matrix $\mathbf{T}$ by means of a scheme similar to Horner's scheme for the division of a polynomial by the binomial $x-a$ (Chapter III). In fact, we find the matrices $\mathbf{D}_i$ determined by the recurrence relations:

$$\left.\begin{aligned}&\mathbf{D}_0=\mathbf{C}_0,\\ &\mathbf{D}_i=\mathbf{C}_i+\mathbf{D}_{i-1}\mathbf{B},\quad i=1,2,\ldots,s,\\ &\mathbf{T}=\mathbf{D}_s.\end{aligned}\right\}\tag{2.39}$$

From the criterion for similarity it follows that if the matrices $\mathbf{A}$ and $\mathbf{B}$ are similar, then their characteristic and minimal polynomials are equal. The converse assertion is not true.

Every matrix $\mathbf{A}$ is similar to its transposed matrix $\mathbf{A}'$, because the matrices $\lambda\mathbf{E}-\mathbf{A}$ and $\lambda\mathbf{E}-\mathbf{A}'$ have the same determinantal factors and therefore are equivalent.

If at least one of the two matrices $\mathbf{A}$ and $\mathbf{B}$, for instance $\mathbf{B}$, is not singular, then the matrices $\mathbf{AB}$ and $\mathbf{BA}$ are similar, because $\mathbf{AB}=\mathbf{B}^{-1}(\mathbf{BA})\mathbf{B}$. This assertion is not true if both the matrices $\mathbf{A}$ and $\mathbf{B}$ are singular.

If the matrices $\mathbf{A}$ and $\mathbf{B}$ are similar and $f(\lambda)$ is any polynomial, then the matrices $f(\mathbf{A})$ and $f(\mathbf{B})$ are similar, and in fact with the same transforming matrix, i.e. if $\mathbf{B}=\mathbf{T}^{-1}\mathbf{AT}$, then $f(\mathbf{B})=\mathbf{T}^{-1}f(\mathbf{A})\mathbf{T}$.

The matrix $\mathbf{A}$ is similar to a matrix $\mathbf{B}$ which is obtained from $\mathbf{A}$ by any interchange of rows and the same interchange of columns. Here $\mathbf{B}=\mathbf{T}^{-1}\mathbf{AT}$, where $\mathbf{T}$ is obtained by applying the same interchange to the columns of the identity matrix.

Any numerical matrix **A** is similar over the field of complex numbers to a triangular (upper or lower, as desired) matrix. In general, a matrix **A**, all of whose elements and characteristic roots lie in a number field K is similar over K to a triangular matrix (subsection 11).

10. The diagonal form of a matrix.

The canonical form of a square matrix **A** of order n, i.e. a matrix **B** of special form, characterizing the class of matrices similar to a given matrix **A**, is not determined uniquely, if we set ourselves the problem of finding the canonical form **B** possibly in a more simple form.

The most simple and important case in practice is the case when the given matrix **A** is similar to a diagonal matrix $\mathbf{B} = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ (p. 72). Here the numbers λ_i lying along the diagonal of the matrix **B** are the characteristic roots of the matrix **A**, taking into account their multiplicity. However, not every matrix **A** is similar to a diagonal matrix. In fact, we have:

First criterion for the existence of a diagonal form. In order that the matrix **A** with elements from a number field K (see Introduction) should be similar over K to the diagonal matrix **B**, i.e. in order that there should exist a non-singular matrix **T** with elements from K, satisfying the equation

$$\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}, \tag{2.40}$$

it is necessary and sufficient that these two conditions should be fulfilled: (a) the last invariant factor $E_n(\lambda)$ (p. 67) or, which is the same thing, the minimal polynomial $\psi(\lambda)$ (p. 78, (2.31)) of the matrix **A** has no multiple roots; (b) all the characteristic numbers of the matrix **A** belong to the field K. Conditions (a) and (b) can be replaced by the equivalent requirement: all the elementary divisors of the matrix **A** over the field K must be polynomials of the first degree ([14], p. 127).

If K is the field of complex numbers, then condition (b) is always fulfilled.

A simpler form is:

A sufficient condition for the existence of a diagonal form. If all the characteristic roots of the matrix **A** with elements from the number field K also lie in the field K and the characteristic polynomial $|\lambda\mathbf{E} - \mathbf{A}|$ does not have multiple roots, then the matrix **A** is similar over K to a diagonal matrix **B**.

Here the characteristic roots of the matrix **A** must lie along the diagonal of the matrix **B**.

By making as small a variation as desired in its elements, any numerical matrix can be transformed into a matrix similar to a diagonal matrix over the field of complex numbers.

Second criterion for the existence of a diagonal form. In order that the matrix **A** with elements from the number field K should be similar over K to a diagonal matrix **B**, it is necessary and sufficient that all the characteristic roots of the matrix **A** should lie in field K and that the matrix **A** should have n linearly independent characteristic vectors.

A rule for constructing a transforming matrix **T**, *reducing the matrix* **A** *to the diagonal form* **B**. (1) If the characteristic polynomial does not have multiple roots, then, taking for each root any characteristic vector $\mathbf{e}_i$, $i = 1, 2, \ldots, n$, we obtain a linearly independent system of vectors (columns). The square matrix **T** formed from these columns is non-singular, and will be the desired transforming matrix.

(2) If the characteristic polynomial has multiple roots, then for every root λ_i of multiplicity k_i we look for a fundamental system of solutions (p. 49) of the homogeneous system of linear equations

$$a_{j1}x_1 + \ldots + (a_{jj} - \lambda_i)x_j + \ldots + a_{jn}x_n = 0, \qquad (2.41)$$

$$j = 1, 2, \ldots, n.$$

If for every root λ_i this fundamental system contains precisely k_i solutions, then, placing the coordinates of the solutions of all these fundamental systems along the columns of the matrix **T**, we obtain a transforming matrix.

(3) If for at least one root λ_i the fundamental system of solutions for the system of equations (2.41) contains less than k_i solutions (we can prove that it cannot contain more than k_i solutions), then the matrix **A** will not be similar to a diagonal matrix, and no transforming matrix **T** exists.

EXAMPLE 1. To show that the matrix

$$\mathbf{A} = \begin{pmatrix} 7 & -12 & 6 \\ 10 & -19 & 10 \\ 12 & -24 & 13 \end{pmatrix}$$

is similar to a diagonal matrix **B**, and to find this matrix **B** and a transforming matrix **T**.

We calculate the characteristic polynomial:

$$\begin{vmatrix} \lambda - 7 & 12 & -6 \\ -10 & \lambda + 19 & -10 \\ -12 & 24 & \lambda - 13 \end{vmatrix}$$

(we add twice the first column to the second column)

$$= \begin{vmatrix} \lambda - 7 & 2\lambda - 2 & -6 \\ -10 & \lambda - 1 & -10 \\ -12 & 0 & \lambda - 13 \end{vmatrix}$$

(we remove the factor $\lambda - 1$ from the second column and we subtract twice the second row from the first)

$$= (\lambda - 1) \begin{vmatrix} \lambda + 13 & 0 & 14 \\ -10 & 1 & -10 \\ -12 & 0 & \lambda - 13 \end{vmatrix}$$

(we expand by the second column)

$$= (\lambda - 1) \begin{vmatrix} \lambda + 13 & 14 \\ -12 & \lambda - 13 \end{vmatrix}$$

(we add the second row to the first, then subtract the first column from the second)

$$= (\lambda - 1) \begin{vmatrix} \lambda + 1 & 0 \\ -12 & \lambda - 1 \end{vmatrix} = (\lambda - 1)^2(\lambda + 1);$$

$$\lambda_1 = \lambda_2 = 1, \quad \lambda_3 = -1.$$

For the root 1 the system (2.41) takes the form

$$\begin{aligned} 6x_1 - 12x_2 + 6x_3 &= 0, \\ 10x_1 - 20x_2 + 10x_3 &= 0, \\ 12x_1 - 24x_2 + 12x_3 &= 0. \end{aligned}$$

Its rank is equal to one and the fundamental system consists of two solutions for instance:

$$\mathbf{e}_1 = (2, 1, 0), \quad \mathbf{e}_2 = (1, 0, -1).$$

For the root -1 we obtain the system

$$\begin{aligned} 8x_1 - 12x_2 + 6x_3 &= 0, \\ 10x_1 - 18x_2 + 10x_3 &= 0, \\ 12x_1 - 24x_2 + 14x_3 &= 0 \end{aligned}$$

of rank 2. The fundamental system consists of one solution, for instance: $\mathbf{e}_3 = (3, 5, 6)$. Because in both cases the number of solutions is equal to the multiplicity of the root, the matrix $\mathbf{A}$ is similar to a diagonal matrix

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Writing the coordinates of the vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ as columns, we obtain the transforming matrix

$$\mathbf{T} = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 0 & 5 \\ 0 & -1 & 6 \end{pmatrix}.$$

The diagonal matrix **B** for the given matrix **A** is defined uniquely to within the order of the diagonal elements. The matrix **T** is not defined uniquely. Thus, in the preceding example we could take as $\mathbf{e}_1$ and $\mathbf{e}_2$ any fundamental system of solutions of the corresponding system of equations. The order of the columns in the matrix **T** corresponds to the order of the diagonal elements of the matrix **B**.

EXAMPLE 2. To determine whether the matrix

$$\mathbf{A} = \begin{pmatrix} 2 & -1 & 2 \\ 5 & -3 & 3 \\ -1 & 0 & -2 \end{pmatrix}.$$

is similar to a diagonal matrix. Calculating the characteristic polynomial, for instance by expanding with respect to the third row, we find

$$|\lambda\mathbf{E} - \mathbf{A}| = (\lambda + 1)^3, \quad \lambda_1 = \lambda_2 = \lambda_3 = -1.$$

We find the characteristic vectors from the system of equations

$$\begin{aligned} 3x_1 - x_2 + 2x_3 &= 0, \\ 5x_1 - 2x_2 + 3x_3 &= 0, \\ -x_1 \qquad - x_3 &= 0. \end{aligned}$$

Its rank is equal to two. The fundamental system consists of one solution, for instance: $\mathbf{e} = (1, 1, -1)$. Because the root -1 of multiplicity three corresponds to one linearly independent characteristic vector, the matrix **A** will not be similar to a diagonal matrix.

If the real matrix **A** of order n is similar over the field of complex numbers to the diagonal matrix

$$\mathbf{B} = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$$

not all of whose diagonal elements are real, then over the field of real numbers it is similar to a real cellular diagonal matrix **B** with cells of the first and second order. In fact if

$$\lambda_1 = a_1 + ib_1, \quad \lambda_2 = a_1 - ib_1, \quad \lambda_3 = a_2 + ib_2,$$
$$\lambda_4 = a_2 - ib_2, \ldots, \quad \lambda_{q2-1} = a_q + ib_q, \quad \lambda_{2q} = a_q - ib_q,$$
$$\lambda_{2q+1} = c_{2q+1}, \ldots, \quad \lambda_n = c_n,$$

where the numbers a_k, b_k, c_k are real and the b_k are positive, then

$$\mathbf{B} = \left\{ \begin{pmatrix} a_1 & -b_1 \\ b_1 & a_1 \end{pmatrix}, \begin{pmatrix} a_2 & -b_2 \\ b_2 & a_2 \end{pmatrix}, \ldots, \begin{pmatrix} a_q & -b_q \\ b_q & a_q \end{pmatrix}, c_{2q+1}, \ldots, c_n \right\}.$$

Thus, the cells of the first order correspond to the real roots, and the cells of the second order to the pairs of conjugate

complex roots of the characteristic polynomial of the matrix **A**. The transforming matrix **T** is constructed in the following way: the system of characteristic vectors of the matrix **A** can be chosen in such a way that if for the complex root $\lambda_{2k-1} = a_k + ib_k$ we take the vector $\mathbf{z}_{2k-1} = \mathbf{x}_k + i\mathbf{y}_k$, where the vectors $\mathbf{x}_k, \mathbf{y}_k$ have real coordinates, then for $\lambda_{2k} = a_k - ib_k$ we take the vector

$$\mathbf{z}_{2k} = \mathbf{x}_k - i\mathbf{y}_k, \qquad k = 1, 2, \ldots, q.$$

For real λ_k we take the vector $\mathbf{z}_k$ with real coordinates. Then along the columns of the matrix **T** we must place the coordinates of the vectors

$$\mathbf{y}_1, \ \mathbf{x}_1, \ \mathbf{y}_2, \mathbf{x}_2, \ldots, \mathbf{y}_q, \ \mathbf{x}_q, \ \mathbf{z}_{2q+1}, \ldots, \mathbf{z}_n$$

([6], pp. 231–2).

11. The Jordan form of a matrix.

Because, by the preceding section, not every matrix is similar to a diagonal matrix, we consider other canonical forms of matrices. One of the simplest canonical forms, which exists for any matrix, is the Jordan form.

A *Jordan cell* is a square matrix of the form

$$\begin{pmatrix} \lambda_0 & 1 & 0 & \ldots & 0 \\ 0 & \lambda_0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & \ldots & \lambda_0 \end{pmatrix}, \tag{2.42}$$

where the same number λ_0 lies along the whole diagonal and all the elements one position above the diagonal are equal to one. Sometimes the series of one's is written not above, but below the diagonal, which makes no essential changes in the theory.

A *Jordan matrix* is a cellular diagonal matrix

$$\mathbf{B} = \{\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_s\}, \tag{2.43}$$

which has any Jordan cells $B_1, B_2, \ldots, B_s$ along its principal diagonal, and all the elements outside these cells are equal to zero.

EXAMPLE 1. The matrix

$$\mathbf{B} = \left(\begin{array}{cc:cc:ccc:c} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

is Jordan. It contains four Jordan cells: two cells of the second order with the number 2 in their diagonals, one cell of the third order with the number 0 in its diagonal and one cell of the first order with the number 0 in its diagonal.

Any square matrix **A**, all of whose elements and characteristic roots lie in the number field K, is similar over K to a Jordan matrix **B**. In particular, any numerical matrix is similar over the field of complex numbers to a Jordan matrix. In general, because any field K is contained in an algebraically closed field $\bar{K}$ (Chapter IV), then any square matrix **A** with elements from K is similar over $\bar{K}$ to a Jordan matrix **B**. For a given matrix **A** the matrix **B** similar to it is defined uniquely to within the order of the arrangement of the Jordan cells $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_s$ along the principal diagonal. This matrix **B** is called the *Jordan form* of the matrix **A**. Thus, every class of similar matrices with elements from an algebraically closed field contains one or a finite number of Jordan matrices, obtained from one another by rearranging the diagonal cells. Two Jordan matrices with elements from a field K are similar to each other if and only if they differ only in the order of sequence of their Jordan cells ([14], p. 124, Theorem 4).

EXAMPLE 2. The matrices

$$\begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

are similar, being Jordan matrices consisting of the same three Jordan cells and differing only in the arrangement of these cells. These matrices are not similar to the matrix

$$\begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix},$$

which contains two Jordan cells.

A rule for constructing the Jordan form of a matrix. If all the elements and characteristic roots of the matrix $\mathbf{A}$ lie in the number field K, then over this field the elementary divisors of the matrix $\mathbf{A}$ or, which is the same thing, of its characteristic matrix $\lambda\mathbf{E} - \mathbf{A}$ will be powers of a linear polynomial, i.e. will have the form

$$(\lambda - \lambda_1)^{k_1}, \quad (\lambda - \lambda_2)^{k_2}, \ldots, (\lambda - \lambda_s)^{k_s}.$$

For every elementary divisor $(\lambda - \lambda_i)^{k_i}$ we construct the Jordan cell

$$\mathbf{B}_i = \begin{pmatrix} \lambda_i & 1 & 0 & \ldots & 0 \\ 0 & \lambda_i & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & \lambda_i \end{pmatrix}$$

of order k_i with the number λ_i in the diagonal. The cellular diagonal matrix $\mathbf{B} = \{\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_s\}$ with these cells along its diagonal will be the Jordan form of the matrix $\mathbf{A}$.

EXAMPLE 3. To find the Jordan form of the matrix

$$\mathbf{A} = \begin{pmatrix} 4 & -5 & 2 & 0 & 0 & 0 \\ 5 & -7 & 3 & 0 & 0 & 0 \\ 6 & -9 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -4 & 4 & 0 \\ 0 & 0 & 0 & -2 & 1 & 2 \end{pmatrix}.$$

We look for the elementary divisors of this matrix. Because it is cellular diagonal, we can look for the elementary divisors of each of its two cells (p. 73). For the cell

$$\mathbf{A}_1 = \begin{pmatrix} 4 & -5 & 2 \\ 5 & -7 & 3 \\ 6 & -9 & 4 \end{pmatrix}$$

the determinantal factors of the characteristic matrix $\lambda\mathbf{E} - \mathbf{A}_1$ will be

$$\mathbf{D}_1(\lambda) = \mathbf{D}_2(\lambda) = 1, \quad \mathbf{D}_3(\lambda) = \lambda^2(\lambda - 1).$$

They can be found by the same method as in Example 2 on p. 69. Therefore the invariant factors are equal to 1, 1, $\lambda^2(\lambda - 1)$ and the elementary divisors are equal to λ^2, $\lambda - 1$. Also for the second cell

$$\mathbf{A}_2 = \begin{pmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{pmatrix}$$

we find the determinantal factors $\mathbf{D}_1(\lambda)=1$, $\mathbf{D}_2(\lambda)=\lambda-2$, $\mathbf{D}_3(\lambda)=(\lambda-2)^3$. The invariant factors are: $1, \lambda-2, (\lambda-2)^2$; the elementary divisors are: $\lambda-2, (\lambda-2)^2$. Therefore the Jordan form of the matrix $\mathbf{A}$ contains four cells and has the form

$$\mathbf{B}=\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

A matrix $\mathbf{T}$ which transforms the given matrix $\mathbf{A}$ to the Jordan form $\mathbf{B}$, can be found by the general method indicated in subsection 9. However, in the given case there is a more economical way of constructing the matrix $\mathbf{T}$, connected with the properties of linear operators and beyond the scope of the present chapter (see, for instance, [6], Chapter 6, section 9, or [20], Problem 1529).

The calculation of the value of the polynomial $f(x)$ in the Jordan matrix $\mathbf{B}=\{\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_s\}$ is carried out by means of the formula $f(\mathbf{B})=\{f(\mathbf{B}_1), f(\mathbf{B}_2), \ldots, f(\mathbf{B}_s)\}$, i.e. it reduces to the calculation of the values of this polynomial in the Jordan cells. The value of the polynomial $f(x)$ in the Jordan cell

$$\mathbf{B}_0=\begin{pmatrix} \lambda_0 & 1 & 0 & \ldots & 0 \\ 0 & \lambda_0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & \lambda_0 \end{pmatrix}$$

of order n is calculated by means of the formula

$$f(\mathbf{B}_0)=\begin{pmatrix} f(\lambda_0) & f'(\lambda_0) & \frac{f''(\lambda_0)}{2!} & \ldots & \frac{f^{(n-1)}(\lambda_0)}{(n-1)!} \\ 0 & f(\lambda_0) & f'(\lambda_0) & \ldots & \frac{f^{(n-2)}(\lambda_0)}{(n-2)!} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & f(\lambda_0) \end{pmatrix}$$

([14], p. 134, (3)). By means of the same formula we can calculate the value of any function $f(x)$ of the Jordan cell $\mathbf{B}_0$, provided only that the value

$$f(\lambda_0), f'(\lambda_0), \ldots, f^{(n-1)}(\lambda_0)$$

has a meaning. In particular, the powers of the Jordan cell $\mathbf{B}_0$ are calculated by means of the formula

$$\mathbf{B}_0^k = \begin{pmatrix} \lambda_0^k & k\lambda_0^{k-1} & C_k^2\lambda_0^{k-2} & \dots & C_k^{n-1}\lambda_0^{k-n+1} \\ 0 & \lambda_0^k & k\lambda_0^{k-1} & \dots & C_k^{n-2}\lambda_0^{n-k+2} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda_0^k \end{pmatrix}.$$

In order to find the Jordan form of the matrix $f(\mathbf{A})$, where $\mathbf{A}$ is a given square matrix and $f(x)$ is a polynomial (or any function for which the value $f(\mathbf{A})$ has a meaning), it is sufficient, by the rule given above for constructing the Jordan form, to know the elementary divisors of the matrix $f(\mathbf{A})$. For their evaluation see [6], Chapter 6, section 7.

12. The natural form and other canonical forms of a matrix.

The Jordan matrices defined in the preceding section do not stand in one–one correspondence to the classes of similar matrices over a given number field K, but only to within permutation of the Jordan cells. We will discuss a canonical form which is free from this defect, but not in so much detail as the Jordan form.

Suppose that we are given a polynomial

$$f(\lambda) = a_0 + a_1\lambda + a_2\lambda^2 + \dots + a_{n-1}\lambda^{n-1} + \lambda^n$$

of degree $n > 0$ with coefficients from a field K with leading coefficient equal to 1. The *companion matrix* of the polynomial $f(\lambda)$ is the matrix

$$\mathbf{B} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \dots & -a_{n-2} & -a_{n-1} \end{pmatrix}.$$

This matrix is square of order equal to the degree n of the polynomial $f(\lambda)$. The last invariant factor of the matrix $\lambda\mathbf{E} - \mathbf{B}$ is equal to $f(\lambda)$, and all the remaining ones are equal to 1.

A matrix of *natural form* (or a matrix having *natural normal form*) is a cellular diagonal matrix

$$\mathbf{B} = \{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_s\},$$

where the diagonal cells $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_s$ are the companion matrices for the polynomials $f_1(\lambda), f_2(\lambda), \ldots, f_s(\lambda)$, each of which (beginning with the second) divides the preceding one. This matrix is a square matrix of order n, equal to the sum of the degrees of the polynomials $f_i(\lambda)$. The last s invariant factors of the matrix $\lambda\mathbf{E}-\mathbf{B}$ are equal respectively to $f_1(\lambda)$, $f_2(\lambda), \ldots, f_s(\lambda)$, and the remaining $n-s$ factors are equal to 1.

Any square matrix $\mathbf{A}$ with elements from some number field K is similar over K to a unique matrix $\mathbf{B}$ of natural form ([14], p. 128, Theorem 5). Two matrices of natural form are similar to each other if and only if they coincide. Thus, every class of similar matrices with elements from a field K contains a unique matrix of natural form.

A rule for constructing the natural form. Let $\mathbf{A}$ be a square matrix of order n with elements from the field K. By one of the methods described in subsection 5, we construct the invariant factors of its characteristic matrix $\lambda\mathbf{E}-\mathbf{A}$. Let

$$f_1(\lambda), f_2(\lambda), \ldots, f_s(\lambda)$$

be all the polynomials of positive degree amongst them, preserving their order (so that each of them divides the preceding one). For each of these polynomials $f_i(\lambda)$ we construct the companion matrix $\mathbf{B}_i$. We note that for polynomials of the first degree $\lambda-a$ the companion matrix will be a matrix of the first order a. The cellular diagonal matrix $\mathbf{B} = \{\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_s\}$ will be the natural form of the matrix $\mathbf{A}$.

EXAMPLE 1. To find the natural form of the matrix $\mathbf{A}$ in Example 3 of the preceding section (p. 90). There we found the elementary divisors of the matrix $\lambda\mathbf{E}-\mathbf{A}$:

$$\lambda^3, \quad \lambda-1, \quad (\lambda-2)^2, \quad \lambda-2.$$

Therefore the invariant factors will be

$$E_6(\lambda) = \lambda^2(\lambda-1)(\lambda-2)^2 = \lambda^5 - 5\lambda^4 + 8\lambda^3 - 4\lambda^2,$$
$$E_5(\lambda) = \lambda - 2.$$

Whence

$$\mathbf{B} = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 4 & -8 & 5 \end{pmatrix}.$$

However, the same matrix $\mathbf{A}$ reduces to the Jordan form with rational elements, like the matrix $\mathbf{A}$ itself.

EXAMPLE 2. To find the natural form of the matrix

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}.$$

The determinantal factors of the matrix

$$\lambda\mathbf{E} - \mathbf{A} = \begin{pmatrix} \lambda - 2 & -1 & 0 \\ 0 & \lambda & -1 \\ 0 & 1 & \lambda \end{pmatrix}$$

will be $D_1(\lambda) = D_2(\lambda) = 1$, because the minor of the second order in the upper right-hand corner is equal to 1, and $D_3(\lambda) = (\lambda - 2)(\lambda^2 + 1)$. The invariant factors are:

$$E_1(\lambda) = E_2(\lambda) = 1, \quad E_3(\lambda) = \lambda^3 - 2\lambda^2 + \lambda - 2.$$

Because the characteristic polynomial has no multiple roots, the matrix **A** is similar to the diagonal matrix

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{pmatrix}$$

over the field of complex numbers. At the same time the matrix **A** is already similar over the field of rational numbers to the matrix of natural form

$$\mathbf{B} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -1 & 2 \end{pmatrix}$$

The quasi-natural form. The natural form is not always the cellular diagonal matrix with the smallest order of diagonal cells amongst the matrices similar to the given matrix **A** over the number field K. This can be seen from Example 1 alone. We will indicate a form which satisfies the condition for the minimal order of diagonal cells for a given field K.

A *quasi-natural matrix* is a cellular diagonal matrix $\mathbf{B} = \{\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_s\}$, where the diagonal cells are the companion matrices for polynomials of the form

$$[\varepsilon_1(\lambda)]^{p_1}, \quad [\varepsilon_2(\lambda)]^{p_2}, \ldots, [\varepsilon_s(\lambda)]^{s_p},$$

where the polynomials $\varepsilon_1(\lambda), \varepsilon_2(\lambda), \ldots, \varepsilon_s(\lambda)$ are irreducible over the field K and have leading coefficients equal to 1.

Any square matrix **A** with elements from some number field K is similar over K to a quasi-natural matrix **B**, defined uniquely to within the order of sequence of the diagonal cells. We will call this matrix **B** the *quasi-natural form* of the matrix **A**. Two quasi-natural matrices with elements from a field K are similar to each other if and only if they differ

in nothing except the order in which their diagonal cells are arranged ([14], p. 129).

A rule for constructing the quasi-natural form. Suppose that we are given a square matrix $\mathbf{A}$ with elements from a field K. We will look for the elementary divisors over K of the characteristic matrix $\lambda\mathbf{E} - \mathbf{A}$ and for each of them we construct its companion matrix. We take these matrices in any order as the diagonal cells of the cellular diagonal matrix $\mathbf{B}$.

EXAMPLE 3. We will find the quasi-natural form of the matrix $\mathbf{A}$ of Example 3 of the preceding section. The elementary divisors of the matrix $\lambda\mathbf{E} - \mathbf{A}$ are equal to λ^2, $\lambda - 1$, $(\lambda - 2)^2$, $\lambda - 2$. Therefore the desired matrix has the form

$$\mathbf{B} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -4 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

This matrix is not a Jordan matrix.

The generalized Jordan form. Examples 1–3 show that the natural and quasi-natural forms are not a generalization of the Jordan form. We will indicate such a form, which can always be found without going outside the limits of a given number field K (and in general of any field K of characteristic zero (Chapter IV)).

A *generalized Jordan cell* for a polynomial of the form $[\varepsilon(\lambda)]^p$, where $\varepsilon(\lambda)$ is irreducible over the number field K and is a polynomial of degree $k > 0$ with leading coefficient equal to 1, and p is a positive integer, is a cellular matrix

$$\mathbf{B}_0 = \begin{pmatrix} \mathbf{A}_0 & \mathbf{E} & 0 & \dots & 0 \\ 0 & \mathbf{A}_0 & \mathbf{E} & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \dots & \mathbf{E} \\ 0 & 0 & 0 & \dots & \mathbf{A}_0 \end{pmatrix},$$

where $\mathbf{A}_0$ is the companion matrix for the polynomial $\varepsilon(\lambda)$ (in the sense indicated at the beginning of this section) and $\mathbf{E}$ is the identity matrix of order k. The matrix $\mathbf{B}_0$ consists of p^2 square cells of order k and has order equal to kp. Its characteristic matrix $\lambda\mathbf{E} - \mathbf{B}_0$ has the unique elementary divisor $[\varepsilon(\lambda)]^p$ over the field K ([14], p. 129, Lemma 4).

A *generalized Jordan matrix* is a cellular diagonal matrix $\mathbf{B} = \{\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_s\}$, where all the diagonal cells are generalized Jordan cells. Any square matrix $\mathbf{A}$ with elements from a number field K (or any field of characteristic zero) is similar over K to a generalized Jordan matrix $\mathbf{B}$. The matrix $\mathbf{B}$ is defined uniquely to within the order of the generalized Jordan cells. It reduces to a Jordan matrix if and only if all the characteristic numbers of the matrix $\mathbf{A}$ lie in the field K [(14], p. 131).

We note that the elementary divisor $[\varepsilon(\lambda)]^p$ of the generalized Jordan cell $\mathbf{B}_0$ is unaltered if we take as the matrix $\mathbf{A}_0$ any matrix of order k with elementary divisor $\varepsilon(\lambda)$. This enables us to find a matrix of another form, similar to the given matrix $\mathbf{A}$. Thus, if K is the field of real numbers, then the only polynomials irreducible over K will be polynomials of the first degree and polynomials of the second degree $\lambda^2 + p\lambda + q$ with complex roots $\alpha \pm i\beta$, $\beta \neq 0$. For the elementary divisor $(\lambda^2 + p\lambda + q)^m$ in the matrix $\mathbf{B}_0$ we can take as $\mathbf{A}_0$ either the associated matrix $\begin{pmatrix} 0 & 1 \\ -q & -p \end{pmatrix}$, or the matrix $\mathbf{C}_0 = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$. In fact,

$$|\lambda\mathbf{E} - \mathbf{C}_0| = \begin{vmatrix} \lambda - \alpha & -\beta \\ \beta & \lambda - \alpha \end{vmatrix} = \lambda^2 - 2\alpha\lambda + \alpha^2 + \beta^2$$

$$= \lambda^2 + p\lambda + q.$$

EXAMPLE 4. To find the generalized Jordan form over the field of real numbers for the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 2 & 3 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 2 & 3 \end{pmatrix}.$$

The characteristic matrix is:

$$\lambda\mathbf{E} - \mathbf{A} = \begin{pmatrix} \lambda - 1 & 1 & 0 & 0 \\ -2 & \lambda - 3 & 1 & 0 \\ 0 & 0 & \lambda - 1 & 1 \\ 0 & 0 & -2 & \lambda - 3 \end{pmatrix}.$$

Expanding its determinant according to Laplace's theorem by the first two columns, we find the determinantal factors $\mathbf{D}_4(\lambda) = (\lambda^2 - 4\lambda + 5)^2$, $\mathbf{D}_3(\lambda) = 1$, because the third order minor in the upper right-hand corner is equal to 1. Therefore the matrix $\lambda\mathbf{E} - \mathbf{A}$ over the field of real numbers has the unique elementary divisor $(\lambda^2 - 4\lambda + 5)^2$. Its roots are $2 \pm i$, both double.

The generalized Jordan form consists of one cell

$$\mathbf{B} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ -5 & 4 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -5 & 4 \end{pmatrix}.$$

If, instead of the companion matrix $\begin{pmatrix} 0 & 1 \\ -5 & 4 \end{pmatrix}$, we take the matrix $\mathbf{C}_0 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, according to the remarks made above we obtain another matrix similar to the matrix **A**, namely:

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}.$$

The matrix **C** (as distinct from **B**) is not a generalized Jordan form of the matrix **A**.

13. Jacobian matrices.

For the calculation of Jacobian determinants, see Chapter 1, section 1, subsection 8. For the proof of the properties of Jacobian matrices mentioned below, see [7], Chapter II, section 1.

A *Jacobian matrix* is a matrix in which all the elements outside the principal diagonal and the two neighbouring parallels are equal to zero. Such a matrix can be written in the form

$$\mathbf{J}_n = \begin{pmatrix} a_1 & -b_1 & 0 & 0 & \dots & 0 \\ -c_1 & a_2 & -b_2 & 0 & \dots & 0 \\ 0 & -c_2 & a_3 & -b_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & a_n \end{pmatrix}. \tag{2.44}$$

We will only consider matrices with real elements, satisfying the conditions

$$b_k c_k > 0, \quad k = 1, 2, \dots, n-1. \tag{2.45}$$

We set

$$\mathbf{D}_0(\lambda) = 1, \quad \mathbf{D}_k(\lambda) = |J_k - \lambda E|, \quad k = 1, 2, \dots, n. \tag{2.46}$$

The properties of the polynomials $\mathbf{D}_k(\lambda)$ (some of these properties are similar to the properties of Sturm series) (see Chapter III):

(1) If $\mathbf{D}_k(\alpha) = 0$, $1 \leqslant k \leqslant n-1$, then $\mathbf{D}_{k-1}(\alpha)$ and $\mathbf{D}_{k+1}(\alpha)$ are distinct from zero and have opposite signs.

(2) As the real variable λ passes through a root of the polynomial $\mathbf{D}_k(\lambda)$ from left to right the product $\mathbf{D}_k(\lambda)\mathbf{D}_{k-1}(\lambda)$, where $k > 0$, changes sign from plus to minus.

(3) All the roots of the polynomial $\mathbf{D}_k(\lambda)$, $k > 0$, are real and distinct.

(4) If a and b, $a < b$, are not roots of the polynomial $\mathbf{D}_k(\lambda)$, then the number of its roots in the segment $[a, b]$ is equal to the increment in the number of changes in sign in the series $\mathbf{D}_0(\lambda), \mathbf{D}_1(\lambda), \ldots, \mathbf{D}_k(\lambda)$ as λ varies from a to b.

(5) Between two neighbouring roots of the polynomial $\mathbf{D}_k(\lambda)$, where $k > 0$, there lies precisely one root of the polynomial $\mathbf{D}_{k-1}(\lambda)$.

(6) If $\lambda_1 < \lambda_2 < \ldots < \lambda_n$ are the roots of the polynomial $\mathbf{D}_n(\lambda)$, then when $\lambda = \lambda_j$ the sequence $\mathbf{D}_0(\lambda), \mathbf{D}_1(\lambda), \ldots, \mathbf{D}_{n-1}(\lambda)$ contains precisely $j - 1$ changes of sign.

(7) If $\mathbf{u}_j = (u_{1j}, u_{2j}, \ldots, u_{nj})$ is a characteristic vector corresponding to the characteristic root of the jth magnitude λ_j of the Jacobian matrix (2.44) with the conditions (2.45), then

$$\mathbf{u}_{kj} = \frac{C_j D_{k-1}(\lambda_j)}{b_1 b_2 \ldots b_{k-1}}, \tag{2.47}$$

where C_j is an arbitrary number distinct from zero, $k = 1, 2, \ldots, n$; $j = 1, 2, \ldots, n$.

A Jacobian matrix is called *normal* if its elements are real and satisfy the conditions $b_k > 0$, $c_k > 0$; $k = 1, 2, \ldots, n-1$. Because the conditions (2.45) are satisfied for a normal Jacobian matrix, it possesses the properties (1)–(7). From this it follows that:

(8) The characteristic numbers λ_i of a normal Jacobian matrix are real and distinct. If $\lambda_1 < \lambda_2 < \ldots < \lambda_n$, then in the series of coordinates of a characteristic vector corresponding to the root λ_j there are precisely $j - 1$ changes of sign.

Let $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ be any n-dimensional vector with real coordinates. The plane figure with vertices $P_1, P_2, \ldots, P_n$, where P_k has the rectangular coordinates $x_k = k$, $y_k = u_k$, $k = 1, 2, \ldots, n$, is called the **u**-*line*. The points of intersection of the **u**-line with the axis Ox are called *nodes of the* **u**-line or **u**-*vectors*.

We consider the vector

$$\mathbf{u}(\lambda) = (u_1(\lambda), u_2(\lambda), \ldots, u_n(\lambda)),$$

whose coordinates are defined by the equations

$$u_1(\lambda) \equiv 1, \quad u_k(\lambda) = \frac{D_{k-1}(\lambda)}{b_1 b_2 \ldots b_{k-1}}, \quad k = 2, 3, \ldots, n. \tag{2.48}$$

The properties of the nodes of a $\mathbf{u}(\lambda)$*-line*:

(9) Every point of a $\mathbf{u}(\lambda)$-line which is common with the axis Ox, when $1 < x < n$, is a node of this line (this follows (from 1)).

(10) a $\mathbf{u}(\lambda)$-line when $\lambda = \lambda_j$ has precisely $j-1$ nodes (this follows from (6)).

(11) The number of nodes of a $\mathbf{u}(\lambda)$-line is equal to the number of roots of the polynomial $\mathbf{D}_{n-1}(\lambda)$ in the interval $(-\infty, \lambda)$ (this follows from (4)).

(12) If $\lambda < \mu$, then between every two nodes of the $\mathbf{u}(\lambda)$-line there lies at least one node of the $\mathbf{u}(\mu)$-line.

(13) As λ increases continuously the nodes of the $\mathbf{u}(\lambda)$-line are moved continuously to the left.

(14) The nodes of two characteristic vectors of a normal Jacobian matrix belonging to two neighbouring characteristic roots occur one after the other.

14. Associated and adjoint matrices.

For associated and reciprocal determinants see Chapter I, section 1, subsection 13. Let $\mathbf{A} = (a_{ij})_1^n$ be a matrix of order $n \geqslant 2$ and let p be an integer, where $1 \leqslant p \leqslant n$. We denote by $N = C_n^p$ the number of sets $i_1 < i_2 < \ldots < i_p$ of p numbers from the n numbers $1, 2, \ldots, n$. We number all these sets in an arbitrary, but henceforward unchanged order:

$$s_1, s_2, \ldots, s_N. \tag{2.49}$$

If s_i is the set $i_1 < i_2 < \ldots < i_p$ and s_j is the set $j_1 < j_2 < \ldots < j_p$ then the minor $\mathbf{A}\begin{pmatrix} i_1 i_2 \ldots i_p \\ j_1 j_2 \ldots j_p \end{pmatrix}$, which stands at the intersection of the p rows with the numbers $i_1, i_2, \ldots, i_p$ and the p columns with the numbers $j_1, j_2, \ldots, j_p$, is denoted by $\tilde{a}_{ij}$.

For a matrix $\mathbf{A}$ of order n the *p-th associated matrix* is the matrix of order $N = C_n^p$, formed from the minors of order p of the matrix $\mathbf{A}$, taken in a definite order

$$\tilde{\mathbf{A}}_p = (\tilde{a}_{ij})_1^N.$$

For instance, $\tilde{\mathbf{A}}_1 = \mathbf{A}$, $\tilde{\mathbf{A}}_n = (|\mathbf{A}|)$.

The properties of associated matrices:

(1) $(\tilde{\mathbf{A}}\tilde{\mathbf{B}})_p = \tilde{\mathbf{A}}_p \cdot \tilde{\mathbf{B}}_p$.

(2) $(\tilde{\mathbf{E}}_n)_p = \mathbf{E}_N$, where $\mathbf{E}_n$ and $\mathbf{E}_N$ are the identity matrices of orders n and N respectively.

(3) If $|\mathbf{A}| \neq 0$, then $(\tilde{\mathbf{A}}^{-1})_p = (\tilde{\mathbf{A}}_p)^{-1}$ ([6], p. 26).

(4) $(\tilde{\mathbf{A}}^k)_p = (\tilde{\mathbf{A}}_p)^k$, where for a non-singular matrix $\mathbf{A}$ this is true for any integer k, and for a singular matrix for any non-negative integer k.

(5) If the matrices $\mathbf{A}$ and $\mathbf{B}$ are similar (p. 81), then their pth associated matrices are similar also, where if $\mathbf{B} = \mathbf{T}^{-1}\mathbf{AT}$, then $\tilde{\mathbf{B}}_p = (\tilde{\mathbf{T}}_p)^{-1} . \tilde{\mathbf{A}}_p . \tilde{\mathbf{T}}_p$.

(6) If $\mathbf{A} = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ is a diagonal matrix, then its pth associated matrix is also diagonal, and $\tilde{\mathbf{A}}_p = \{\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_N\}$, where along the principal diagonal of the matrix $\tilde{\mathbf{A}}_p$ there stand all possible products of p diagonal elements of the matrix $\mathbf{A}$ in the following order: if s_i is the set $i_1 < i_2 < \ldots < i_p$, then $\tilde{\lambda}_i = \lambda_{i1}\lambda_{i2} \ldots \lambda_{ip}$.

(7) As the numbering of the sets in the sequence (2.49) is changed, an identical rearrangement of the rows and columns is made in the pth associated matrix $\tilde{\mathbf{A}}_p$, i.e. the matrix $\tilde{\mathbf{A}}_p$ is changed to a similar matrix (p. 83), and the characteristic roots of this matrix are unchanged. Therefore to calculate the characteristic roots of the pth associated matrix one can choose the most convenient order of numbering the sequence (2.49).

(8) If the matrix $\mathbf{A}$ is triangular and the sets in the sequence (2.49) are numbered in lexicographical order, then the pth associated matrix $\tilde{\mathbf{A}}_p$ will also be triangular. Here by lexicographical order we mean the arrangement in which the set $i_1 < i_2 < \ldots < i_p$ precedes the set $j_1 < j_2 < \ldots < j_p$, if the first of the differences $j_1 - i_1, j_2 - i_2, \ldots, j_p - i_p$ distinct from zero is positive. The diagonal elements of the triangular matrices $\mathbf{A}$ and $\tilde{\mathbf{A}}_p$ are connected in the same way as was indicated in property (6).

(9) The characteristic roots of the pth associated matrix $\tilde{\mathbf{A}}_p$ (taking into account their multiplicity) are equal to all possible products of p of the characteristic roots of the matrix $\mathbf{A}$.

(10) If the matrix $\mathbf{A}$ is similar to the diagonal matrix $\mathbf{B} = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, then the pth associated matrix $\tilde{\mathbf{A}}_p$ is similar to the diagonal matrix $\tilde{\mathbf{B}}_p = \{\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_N\}$, where $\tilde{\lambda}_i$ is obtained by the rule given in property (6).

If $\mathbf{T}$ is a matrix which transforms $\mathbf{A}$ into $\mathbf{B}$, then $\tilde{\mathbf{T}}_p$ will be a matrix which transforms $\tilde{\mathbf{A}}_p$ into $\tilde{\mathbf{B}}_p$. In other words, if the matrix $\mathbf{A}$ has n linearly independent characteristic vectors and from their coordinates (written in columns) we form the matrix $\mathbf{T}$, then the pth associated matrix $\tilde{\mathbf{A}}_p$

has N linearly independent characteristic vectors, whose coordinates form the columns of the matrix $\tilde{\mathbf{T}}_p$. This means that if s_i is the set $i_1 < i_2 < \ldots < i_p$ and s_j is the set $j_1 < j_2 < \ldots < j_p$ in the sequence (2.49), then the ith coordinate of the characteristic vector of the matrix $\tilde{\mathbf{A}}_p$, corresponding to the characteristic root $\tilde{\lambda}_j = \lambda_{j1}\ \lambda_{j2} \ldots \lambda_{jp}$ is equal to the minor $\mathbf{T}\begin{pmatrix} i_1\, i_2 \ldots i_p \\ j_1 j_2 \ldots j_p \end{pmatrix}$ of order p in the matrix $\mathbf{T}$ ([6], pp. 67–68, or [7], p. 81).

The *associate* of the matrix $\mathbf{A} = (a_{ij})_1^n$ is the matrix $\tilde{\mathbf{A}} = (\tilde{a}_{ij})_1^n$, where $\tilde{a}_{ij}$ is the minor of the element a_{ij} of the matrix $\mathbf{A}$. Thus, $\tilde{\mathbf{A}}$ is a special form of $\tilde{\mathbf{A}}_{n-1}$. Then:

(11) $\tilde{\tilde{\mathbf{A}}} = |\mathbf{A}|^{n-2}\mathbf{A}$, where when $n = 2$ it is necessary to assume that $|\mathbf{A}|^0 = \mathbf{1}$ ([20], Problem 968).

(12) If the matrix $\mathbf{A}$ is orthogonal (or unitary, subsection 16), then the pth associated matrix $\tilde{\mathbf{A}}_p$ is orthogonal (respectively, unitary) ([20], Problem 973, 974).

The *adjoint* (or *reciprocal*) matrix of the matrix $\mathbf{A} = (a_{ij})_1^n$ is the matrix $|\hat{\mathbf{A}}| = (\hat{a}_{ij})_1^n$, where $\hat{a}_{ij}$ is the cofactor of the element a_{ij} in the matrix $\mathbf{A}$. It has the following properties:

(13) $\mathbf{A}|\hat{\mathbf{A}}| = |\hat{\mathbf{A}}|\,\mathbf{A} = |\mathbf{A}|\mathbf{E}$

(14) $\hat{\hat{\mathbf{A}}} = |\mathbf{A}|^{n-2}\mathbf{A}$ ([20], Problem 966).

(15) $\widehat{\mathbf{AB}} = \hat{\mathbf{B}}\hat{\mathbf{A}}$ ([20], Problem 967).

For the matrix adjoint to the characteristic matrix $\lambda\mathbf{E} - \mathbf{A}$, see p. 78.

15. Kronecker product of matrices

For the Kronecker product of determinants see Chapter I, section 1, subsection 15. Suppose that all the pairs (i, j), where $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$, be numbered in an arbitrary, but henceforward unchanged order:

$$s_1, s_2, \ldots, s_{mn}. \tag{2.50}$$

The *Kronecker* (or *direct*) product of two matrices $\mathbf{A} = (a_{ij})_1^m$ and $\mathbf{B} = (b_{ij})_1^n$ is the matrix $\mathbf{C} = (c_{ij})_1^{mn}$, consisting of all possible products of elements of the matrices $\mathbf{A}$ and $\mathbf{B}$ in the proper order. In fact, $c_{ij} = a_{i_1 j_1} b_{i_2 j_2}$, where $(i_1, i_2) = s_i$, $(j_1, j_2) = s_j$, $i, j = 1, 2, \ldots, mn$. It is denoted by $\mathbf{C} = \mathbf{A} \times \mathbf{B}$.

The properties of the Kronecker product of matrices.

(1) In the general case the question of the commutativity and associativity of the Kronecker product has no meaning, because the way in which the sequence (2.50) is given only enables one to multiply a matrix of order m by a matrix of order n.

(2) The Kronecker product is distributive:

$$(\mathbf{A}+\mathbf{B})\times\mathbf{C}=\mathbf{A}\times\mathbf{C}+\mathbf{B}\times\mathbf{C},$$

where $\mathbf{A}$ and $\mathbf{B}$ are of order m, and $\mathbf{C}$ is of order n;

$$\mathbf{A}\times(\mathbf{B}+\mathbf{C})=\mathbf{A}\times\mathbf{B}+\mathbf{A}\times\mathbf{C},$$

where $\mathbf{A}$ is of order m, and $\mathbf{B}$ and $\mathbf{C}$ are of order n.

(3) $(\mathbf{AB})\times(\mathbf{CD})=(\mathbf{A}\times\mathbf{C})(\mathbf{B}\times\mathbf{D})$, where $\mathbf{AB}$, are of order m and $\mathbf{C}$, $\mathbf{D}$ are of order n ([20], Problem 963).

More often than the general concept of a Kronecker product one applies the following two special forms of it.

The *right Kronecker product* of the matrices $\mathbf{A}=(a_{ij})_1^m$ and $\mathbf{B}=(b_{ij})_1^n$ is the partitioned matrix $\mathbf{A}\cdot\mathbf{B}=\mathbf{C}=(\mathbf{C}_{ij})_1^m$, where $\mathbf{C}_{ij}=a_{ij}\mathbf{B}$, $i, j=1, 2, \ldots, m$.

It is obtained from the general Kronecker product if the pairs (i, j) in the sequence (2.50) are numbered in lexicographical order:

$$(1, 1), (1, 2), \ldots, (1, n),$$

$$(2, 1), (2, 2), \ldots, (2, n), \ldots, (m, 1), (m, 2), \ldots, (m, n).$$

The *left Kronecker product* of the same matrices is the partitioned matrix $\mathbf{A}.\times\mathbf{B}=\mathbf{D}=(\mathbf{D}_{ij})_1^n$, where $\mathbf{D}_{ij}=\mathbf{A}.\,b_{ij}$ $(i, j=1, 2, \ldots, n)$.

It is obtained if in the sequence (2.50) the pairs (i, j), read from right to left, are numbered in lexicographical order:

$$(1, 1), (2, 1), \ldots, (m, 1),$$

$$(1, 2), (2, 2), \ldots, (m, 2), \ldots, (1, n), (2, n), \ldots, (m, n).$$

These products have a meaning for matrices of any order and, in addition to properties (2) and (3), they also possess the following:

(4) $\mathbf{A}\times.\mathbf{B}=\mathbf{B}.\times\mathbf{A}$.

(5) Associativity:

$$\mathbf{A}\times.(\mathbf{B}\times.\mathbf{C})=(\mathbf{A}\times.\mathbf{B})\times\mathbf{C},\quad \mathbf{A}.\times(\mathbf{B}.\times\mathbf{C})$$
$$=(\mathbf{A}.\times\mathbf{B}).\times\mathbf{C}.$$

(6) If $\mathbf{E}_k$ is the identity matrix of order k, then

$$\mathbf{E}_m \times . \mathbf{E}_n = \mathbf{E}_{m\cdot} \times \mathbf{E}_{n\cdot} = \mathbf{E}_{mn}.$$

(7) If $\mathbf{A}$ and $\mathbf{B}$ are non-singular matrices, then $(\mathbf{A} \times . \mathbf{B})^{-1} = \mathbf{A}^{-1} \times \mathbf{B}^{-1}$, $(\mathbf{A}. \times \mathbf{B})^{-1} = \mathbf{A}^{-1}. \times \mathbf{B}^{-1}$ ([20], Problem 964).

(8) If the matrices $\mathbf{A}_1, \mathbf{B}_1$ are similar, respectively, to the matrices $\mathbf{A}_2, \mathbf{B}_2$, then the Kronecker products $\mathbf{A}_1 \times \mathbf{B}_1$ and $\mathbf{A}_2 \times \mathbf{B}_2$ (taken for any two arrangements of the pairs of indices in the sequence (2.50)) are similar to each other. In particular, if the order of numbering the pairs in the sequence (2.50) is changed, the matrix $\mathbf{A} \times \mathbf{B}$ is replaced by a similar matrix ([20], Problem 1056).

(9) The Kronecker product of diagonal matrices is diagonal.

(10) The right and left Kronecker products of two upper (or two lower) triangular matrices is an upper (lower) triangular matrix.

(11) If $\lambda_1, \lambda_2, \ldots, \lambda_m$ is the spectrum of the matrix $\mathbf{A}$ and $\mu_1, \mu_2, \ldots, \mu_n$ is the spectrum of the matrix $\mathbf{B}$, then the spectrum of their Kronecker product $\mathbf{A} \times \mathbf{B}$ consists of all possible products $\lambda_i \cdot \mu_j$, $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$ ([20], Problem 1146).

(12) If the matrices $\mathbf{A}$ and $\mathbf{B}$ are similar diagonal matrices, then this is also true for their Kronecker product $\mathbf{A} + \mathbf{B}$.

16. Orthogonal and unitary matrices.

The *scalar product* of two n-dimensional vectors $\mathbf{a} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, $\mathbf{b} = (\beta_1, \beta_2, \ldots, \beta_n)$ in the case of real coordinates is the number

$$(\mathbf{a}, \mathbf{b}) = \alpha_1\beta_1 + \alpha_2\beta_2 + \ldots + \alpha_n\beta_n, \tag{2.51}$$

and in the case of complex coordinates is the number

$$(\mathbf{a}, \mathbf{b}) = \alpha_1\overline{\beta}_1 + \alpha_2\overline{\beta}_2 + \ldots + \alpha_n\overline{\beta}_n, \tag{2.52}$$

where the bars denote the numbers which are the complex conjugates of the given numbers. We can give another, axiomatic, definition of the scalar product (see [11], pp. 211–12).

If $(\mathbf{a}, \mathbf{b}) = 0$, then the vectors $\mathbf{a}$ and $\mathbf{b}$ are said to be *orthogonal*. If $(\mathbf{a}, \mathbf{a}) = 1$, then the vector $\mathbf{a}$ is called a *unit* or a *normed vector*. The system of vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_s$ is called *orthonormal* if

$$(\mathbf{e}_i, \mathbf{e}_j) = \delta_{ij} = \begin{cases} 1 & \text{when } i = j; \\ 0 & \text{when } i \neq j; \end{cases} \quad i, j = 1, 2, \ldots, s. \tag{2.53}$$

A square matrix $\mathbf{A}$ is called *orthogonal* if its elements are real and satisfy one of the following five equivalent conditions:

(a) the rows form an ortho-normal system, (b) the columns form an ortho-normal system, (c) $\mathbf{AA'} = \mathbf{E}$, (d) $\mathbf{A'A} = \mathbf{E}$, (e) $\mathbf{A'} = \mathbf{A}^{-1}$. Here $\mathbf{A'}$ signifies the transposed matrix $\mathbf{A}$. In a similar way we can also define a complex orthogonal matrix, but we will limit ourselves to real matrices.

In the case of complex numbers a similar concept is the concept of a unitary matrix. The matrix conjugate to the matrix $\mathbf{A}$ is the matrix $\mathbf{A^*} = \overline{\mathbf{A}}'$, obtained from $\mathbf{A}$ by transposing it and replacing all its elements by their complex conjugates.

A square matrix $\mathbf{A}$ with any complex (in particular, real) elements is called *unitary*, if one of the following five equivalent conditions is satisfied: (a) the rows from an ortho-normal system (here the scalar product is defined by the formula (2.52)), (b) the columns form an ortho-normal system, (c) $\mathbf{AA^*} = \mathbf{E}$, (d) $\mathbf{A^*A} = \mathbf{E}$, (e) $\mathbf{A^*} = \mathbf{A}^{-1}$.

A unitary matrix is orthogonal if and only if it is real. An orthogonal or unitary matrix is not singular, and, moreover, the determinant of an orthogonal matrix is equal to ± 1, and the determinant of a unitary matrix is equal in modulus to 1.

The product of two orthogonal matrices, and the matrix inverse to an orthogonal matrix, are orthogonal. In other words, all the orthogonal matrices of a given order form a group with respect to multiplication (Chapter IV). Then the same is true for unitary matrices.

An orthogonal matrix may not have real characteristic roots. For instance, the matrix $\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$, where $\varphi \neq k\pi$, has the characteristic roots $\cos\varphi \pm i \sin\varphi$, and $\sin\varphi \neq 0$. The characteristic roots of a unitary (in particular, an orthogonal) matrix have modulus equal to 1. Therefore all the real characteristic roots (if they exist) are equal to ± 1. Characteristic vectors of an orthogonal or unitary matrix belonging to two distinct characteristic roots are orthogonal.

Two matrices $\mathbf{A}$ and $\mathbf{B}$ are called *orthogonally* (or *unitarily*) *similar* if there exists an orthogonal (respectively, a unitary) matrix $\mathbf{T}$, satisfying the equation

$$\mathbf{B} = \mathbf{T}^{-1}\mathbf{AT}. \tag{2.54}$$

A matrix $\mathbf{A}$ is unitary if and only if it is unitarily similar to a diagonal matrix $\mathbf{B}$ whose diagonal elements have modulus equal to 1, i.e.

$$\mathbf{B} = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}, \ |\lambda_i| = 1, \quad i = 1, 2, \ldots, n. \tag{2.55}$$

Here the numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$, which lie along the diagonal of the matrix **B**, form the spectrum of the matrix **A**, and the columns of the transforming matrix **T** consist of the coordinates of an ortho-normal system of n characteristic vectors of the matrix **A**, belonging to the characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ in the corresponding order ([6], p. 224, Theorem 6′).

A real matrix **A** of order n is orthogonal if and only if it is orthogonally similar to a canonical orthogonal matrix

$$\mathbf{B} = \left\{ \begin{pmatrix} \cos\varphi_1 & -\sin\varphi_1 \\ \sin\varphi_1 & \cos\varphi_1 \end{pmatrix}, \ldots \right.$$

$$\left. \ldots, \begin{pmatrix} \cos\varphi_q & -\sin\varphi_q \\ \sin\varphi_q & \cos\varphi_q \end{pmatrix}, 1, \ldots, 1, -1, \ldots, -1 \right\}. \qquad (2.56)$$

Here $\sin\varphi_k > 0$, $k = 1, 2, \ldots, q$; the number of elements $+1$ is equal to r, the number of elements -1 is equal to s; $2q + r + s = n$. Some of the numbers q, r, s may be equal to zero, i.e. cells of some of the given types may not occur ([6], p. 233). Cells of the second order correspond to pairs of complex conjugate characteristic numbers, and of the first order to real characteristic numbers of the matrix **A**, so that the spectrum of the matrix **A** has the form

$$\lambda_1 = \cos\varphi_1 + i\sin\varphi_1, \quad \lambda_2 = \cos\varphi_1 - i\sin\varphi_1, \ldots$$

$$\ldots, \lambda_{2q-1} = \cos\varphi_q + i\sin\varphi_q, \quad \lambda_{2q} = \cos\varphi_q - i\sin\varphi_q,$$

$$\lambda_{2q+1} = \ldots = \lambda_{2q+r} = 1, \quad \lambda_{2q+r+1} = \ldots = \lambda_n = -1.$$

An orthogonal matrix **T** which transforms the matrix **A** into the canonical form (2.56) is constructed in the following way. For the complex characteristic root $\lambda_{2k-1} = \cos\varphi_k + i\sin\varphi_k$, $k = 1, 2, \ldots, q$, of multiplicity p there exist p orthogonal characteristic vectors $\mathbf{z}_1, \ldots, \mathbf{z}_p$ with complex coordinates. They can be found as the solution of a system of homogeneous linear equations with matrix $\mathbf{A} - \lambda_{2k-1}\mathbf{E}$, where to find the vector $\mathbf{z}_s$ it is necessary to require that it be orthogonal to the vectors $\mathbf{z}_1, \ldots, \mathbf{z}_{s-1}$ already constructed. If $\mathbf{z}_s = \mathbf{x}_s + \mathbf{y}_s$ $(s = 1, \ldots, p)$, where $\mathbf{x}_s, \mathbf{y}_s$ are vectors with real coordinates, then the vectors $\mathbf{x}_s$ and $\mathbf{y}_s$ turn out to be orthogonal and have the same length ([20], Problem 1569). By norming these vectors, we obtain two vectors whose coordinates must be placed in two neighbouring columns of the matrix **T**. The numbers of these columns must coincide with the numbers

of the columns of the sth of the p identical diagonal cells of the form $\begin{pmatrix} \cos\varphi_k & -\sin\varphi_k \\ \sin\varphi_k & \cos\varphi_k \end{pmatrix}$ of the canonical matrix (2.56). Here it is necessary to write first the column of the coordinates of the vector obtained by norming the coefficients of the imaginary part, i.e. $\mathbf{y}_s$, and then of the real part $\mathbf{x}_s$. In the opposite order the cells of the matrix (2.56) would have the form $\begin{pmatrix} \cos\varphi_k & \sin\varphi_k \\ -\sin\varphi_k & \cos\varphi_k \end{pmatrix}$.

For a real characteristic root λ_k of multiplicity p in the same way we can find an ortho-normal system of p real solutions of a system of homogeneous linear equations with matrix $\mathbf{A} - \lambda_k\mathbf{E}$. The coordinates of these vectors must be placed in the p columns of the matrix $\mathbf{T}$ with the same numbers as the columns of the matrix $\mathbf{B}$ containing λ_k in the diagonal.

EXAMPLE. For the matrix

$$\mathbf{A} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{pmatrix}$$

to find an orthogonal matrix $\mathbf{T}$ which reduces $\mathbf{A}$ to a canonical orthogonal matrix $\mathbf{B}$.

By calculating the scalar products of the rows of the matrix $\mathbf{A}$, it is easy to verify that these rows are ortho-normal. This means that the matrix $\mathbf{A}$ is orthogonal, and the problem is soluble. We will calculate the characteristic numbers of the matrix $\mathbf{A}$:

$$|\mathbf{A} - \lambda\mathbf{E}| = \begin{vmatrix} \frac{2}{3} - \lambda & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} - \lambda & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} - \lambda \end{vmatrix}$$

(we add the second and third rows to the first)

$$= \begin{vmatrix} 1-\lambda & 1-\lambda & 1-\lambda \\ \frac{2}{3} & \frac{2}{3} - \lambda & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} - \lambda \end{vmatrix}$$

(subtract the first column from the remaining ones and expand the determinant by the first row)

$$= (1-\lambda)\begin{vmatrix} -\lambda & -1 \\ 1 & 1-\lambda \end{vmatrix} = (1-\lambda)(\lambda^2 - \lambda + 1);$$

$$\lambda_1 = \frac{1}{2} + \frac{1}{2} i\sqrt{3} = \cos\frac{\pi}{3} + i\sin\frac{\pi}{3},$$

$$\lambda_2 = \frac{1}{2} - \frac{1}{2} i\sqrt{3} = \cos\frac{\pi}{3} - i\sin\frac{\pi}{3}, \quad \lambda_3 = 1.$$

The canonical orthogonal matrix orthogonally similar to the matrix **A** has the form

$$\mathbf{B} = \begin{pmatrix} \cos\frac{\pi}{3} & -\sin\frac{\pi}{3} & 0 \\ \sin\frac{\pi}{3} & \cos\frac{\pi}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We will find an orthogonal transforming matrix **T** which satisfies the equation (2.54). The coordinates of a characteristic vector of the matrix **A** for the value $\lambda_1 = \frac{1}{2} + \frac{1}{2} i\sqrt{3}$ can be found from the system of equations

$$\left(\frac{1}{6} - \frac{1}{2} i\sqrt{3}\right) x_1 - \frac{1}{3} x_2 + \frac{2}{3} x_3 = 0,$$

$$\frac{2}{3} x_1 + \left(\frac{1}{6} - \frac{1}{2} i\sqrt{3}\right) x_2 - \frac{1}{3} x_3 = 0,$$

$$-\frac{1}{3} x_1 + \frac{2}{3} x_2 + \left(\frac{1}{6} - \frac{1}{2} i\sqrt{3}\right) x_3 = 0.$$

Because the determinant of this system must be equal to zero and the equations are not proportional, the rank of the system is two and the solution is defined uniquely to within a numerical factor. Setting $x = 1$, we find from the first two equations that

$$x_2 = -\frac{1}{2} - \frac{1}{2} i\sqrt{3}, \quad x_3 = -\frac{1}{2} + \frac{1}{2} i\sqrt{3}.$$

Separating the real part and the coefficients for the imaginary part, we obtain two vectors:

$$f_1 = \left(0, -\frac{1}{2}\sqrt{3}, \frac{1}{2}\sqrt{3}\right), \quad f_2 = \left(1, -\frac{1}{2}, -\frac{1}{2}\right),$$

which are orthogonal and have identical length. Norming these vectors, we find

$$e_1 = \left(0, -\frac{1}{2}\sqrt{2}, \frac{1}{2}\sqrt{2}\right), \quad e_2 = \left(\frac{1}{3}\sqrt{6}, -\frac{1}{6}\sqrt{6}, -\frac{1}{6}\sqrt{6}\right).$$

The coordinates of a characteristic vector of the matrix **A** for the value $\lambda_3 = 1$ can be found from the system of equations

$$-\frac{1}{3}x_1 - \frac{1}{3}x_2 + \frac{2}{3}x_3 = 0,$$

$$\frac{2}{3}x_1 - \frac{1}{3}x_2 - \frac{1}{3}x_3 = 0,$$

$$-\frac{1}{3}x_1 + \frac{2}{3}x_2 - \frac{1}{3}x_3 = 0.$$

Subtracting the first equation from the remaining ones we find: $x_1 - x_3 = 0$, $x_2 - x_3 = 0$, whence $x_1 = x_2 = x_3$. Because the solution is determined to within a factor, we can set $x_1 = x_2 = x_3 = 1$. Hence $\mathbf{f}_3 = (1, 1, 1)$. Norming, we find

$$\mathbf{e}_3 = \left(\frac{1}{3}\sqrt{3},\ \frac{1}{3}\sqrt{3},\ \frac{1}{3}\sqrt{3}\right).$$

Writing the coordinates of the vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ in columns, we obtain the transforming matrix

$$\mathbf{T} = \begin{pmatrix} 0 & \frac{1}{3}\sqrt{6} & \frac{1}{3}\sqrt{3} \\ -\frac{1}{2}\sqrt{2} & -\frac{1}{6}\sqrt{6} & \frac{1}{3}\sqrt{3} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{6}\sqrt{6} & \frac{1}{3}\sqrt{3} \end{pmatrix}.$$

Remark. The solution of the preceding problem enables one to interpret the geometrical significance of the matrix **A**. This matrix gives on an ortho-normal basis an orthogonal transformation of three-dimensional Euclidean space R ([11], p. 221). The canonical form **B** shows that this transformation is equivalent to rotating the space R about the axis determined by the vector $\mathbf{f}_3$ through an angle $\pi/3$ in the direction from $\mathbf{e}_1$ to $\mathbf{e}_2$. Because $|\mathbf{T}| = 1$, the orientation of the points $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ coincides with the orientation of the initial basis. In the case of three-dimensional space, the geometrical meaning of the transformation given by the matrix **A** can be explained more simply without finding the whole of the matrix **T**. In this case the characteristic polynomial has at least one real root, in our example $\lambda_3 = 1$. The corresponding characteristic vector $\mathbf{f}_3 = (1, 1, 1)$ defines the axis of rotation. Let $\mathbf{A} = (a_{ij})_1^3$. The similar matrices **A** and **B** have the same characteristic polynomial, in particular they have the same trace (p. 76). Hence

$$2\cos\varphi + 1 = a_{11} + a_{22} + a_{33}, \quad \cos\varphi = \frac{a_{11} + a_{22} + a_{33} - 1}{2}.$$

In our example this gives $\cos\varphi = \frac{1}{2}$, $\varphi = \pi/3$. It remains to find the direction of rotation. For this it is sufficient to take any vector **a**, not collinear with $\mathbf{f}_3$, the vector **b**, obtained from **a** under the given transformation, and the vector $\mathbf{f}_3$. If we look from the end of the vector $\mathbf{f}_3$ towards the beginning, then the direction of rotation will be from **a** to **b**. If we calculate the determinant of the coordinates of the vectors **a**, **b**, $\mathbf{f}_3$, we discover whether the

orientations of the original basis and the points $\mathbf{a}, \mathbf{b}, \mathbf{f}_3$ coincide, and in this way we determine the direction of rotation with respect to the original basis. In our example we can take the vector $\mathbf{a} = (1, 0, 0)$. The column of the coordinates of its image $\mathbf{b}$ is equal to the product of the matrix $\mathbf{A}$ by the column of coordinates of the vector $\mathbf{a}$, i.e. is equal to the first column of the matrix $\mathbf{A}$. Thus $\mathbf{b} = (\frac{2}{3}, \frac{2}{3}, -\frac{1}{3})$. We calculate the determinant whose columns consist of the coordinates of the vectors $\mathbf{a}, \mathbf{b}, \mathbf{f}_3$.

$$\begin{vmatrix} 1 & \frac{2}{3} & 1 \\ 0 & \frac{2}{3} & 1 \\ 0 & -\frac{1}{3} & 1 \end{vmatrix} = 1 > 0.$$

This means that the transformation given by the matrix $\mathbf{A}$ is a rotation about the axis which is determined by the vector $\mathbf{f}_3 = (1, 1, 1)$, through an angle $\pi/3$ in the direction coinciding with a rotation from the first axis of the original system of coordinates to the second, if we look along the plane of these axes on the positive side of the third semi-axis. This coincides with the rotation obtained earlier through an angle $\pi/3$ from $\mathbf{e}_1$ to $\mathbf{e}_2$.

Schur's theorem. Any square numerical matrix is unitarily similar to a triangular matrix (upper or lower, as desired) ([20], Problem 1548).

Any square numerical matrix $\mathbf{A}$ can be represented both in the form $\mathbf{A} = \mathbf{UC}$, and in the form $\mathbf{A} = \mathbf{DV}$, where $\mathbf{C}$, $\mathbf{D}$ are upper (lower) triangular and $\mathbf{U}$, $\mathbf{V}$ are unitary matrices. If the matrix $\mathbf{A}$ is not singular, then both these decompositions are unique to within a diagonal unitary factor $\mathbf{M} = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$, $|\varepsilon_i| = 1$. If the matrix $\mathbf{A}$ is real, then the matrices $\mathbf{C}$, $\mathbf{D}$, $\mathbf{U}$, $\mathbf{V}$ are also real. Thus, any real matrix can be decomposed into the product of orthogonal and triangular matrices.

17. Symmetric, skew-symmetric and Hermitian matrices.

For the rank of symmetric and skew-symmetric matrices, see Chapter I, section 2, subsection 4.

Let $\mathbf{A}'$ be the transposed matrix for $\mathbf{A}$, and $\mathbf{A}^* = \overline{\mathbf{A}}'$ be the matrix which is conjugate to $\mathbf{A}$. If $\mathbf{A}$ is real, then $\mathbf{A}^* = \mathbf{A}'$. (In the present section we will only consider square matrices.)

A matrix $\mathbf{A}$ is called *symmetric* if $\mathbf{A}' = \mathbf{A}$, *skew-symmetric* if $\mathbf{A}' = -\mathbf{A}$, *Hermitian* if $\mathbf{A}^* = \mathbf{A}$, and *skew-Hermitian* if $\mathbf{A}^* = -\mathbf{A}$. Real Hermitian matrices coincide with symmetric matrices, and real skew-Hermitian with skew-symmetric.

For complex orthogonal, symmetric and skew-symmetric matrices see [6], Chapter 11.

Hermitian and skew-Hermitian matrices are connected as follows. If in the relation $\mathbf{A} = \pm i\mathbf{B}$ one of the two matrices is Hermitian, then the other is skew-Hermitian, and conversely.

Any matrix $\mathbf{A}$ can be represented uniquely in the form of a sum $\mathbf{A} = \mathbf{B} + \mathbf{C}$, where $\mathbf{B}$ is a symmetric and $\mathbf{C}$ a skew-symmetric matrix. Here

$$\mathbf{B} = \frac{\mathbf{A} + \mathbf{A}'}{2}, \quad \mathbf{C} = \frac{\mathbf{A} - \mathbf{A}'}{2}.$$

Further, any matrix $\mathbf{A}$ can be represented uniquely in the form $\mathbf{A} = \mathbf{B} + \mathbf{C}$, where $\mathbf{B}$ is Hermitian and $\mathbf{C}$ is skew-Hermitian. Here

$$\mathbf{B} = \frac{\mathbf{A} + \mathbf{A}^*}{2}, \quad \mathbf{C} = \frac{\mathbf{A} - \mathbf{A}^*}{2}.$$

Finally, any matrix $\mathbf{A}$ can be represented uniquely in the form

$$\mathbf{A} = \mathbf{B} + i\mathbf{C}, \tag{2.57}$$

where both the matrices $\mathbf{B}$, $\mathbf{C}$ are Hermitian. Here

$$\mathbf{B} = \frac{\mathbf{A} + \mathbf{A}^*}{2}, \quad \mathbf{C} = \frac{\mathbf{A} - \mathbf{A}^*}{2i}.$$

The decomposition (2.57) is the matrix analogue of the representation of the complex number a in the form $a = b + ic$ with real numbers b, c.

From the rule for transposing the product $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$ and the rule for replacing a number by its conjugate $\overline{\mathbf{AB}} = \overline{\mathbf{A}}\,\overline{\mathbf{B}}$, there follows the equation

$$(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*. \tag{2.58}$$

From this it is easy to deduce that the product of two symmetric (or Hermitian) matrices $\mathbf{A}$ and $\mathbf{B}$ will be symmetric (Hermitian) if and only if $\mathbf{AB} = \mathbf{BA}$. Further, the product of two skew-symmetric (or skew-Hermitian) matrices $\mathbf{A}$ and $\mathbf{B}$ will be skew-symmetric, (skew-Hermitian) if and only if $\mathbf{AB} = -\mathbf{BA}$.

Any matrix can be represented in the form of a product of two symmetric matrices, one of which is not singular (see [24], Problem 978, or [20], Problem 1144).

The characteristic roots of a Hermitian (in particular, a real symmetric) matrix are real. The characteristic roots of a skew-Hermitian (in particular, a real skew-symmetric)

matrix are purely imaginary, i.e. have the form bi, where b is a real number (in particular, some of these numbers may be equal to zero). The characteristic vectors of a matrix of any of the given forms which belong to distinct characteristic roots are orthogonal.

A matrix **A** is Hermitian if and only if it is unitarily similar to a real diagonal matrix $\mathbf{B} = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ ([6], p. 223, Theorem 5'). Here the numbers λ_i which lie along the diagonal of the matrix **B** form the spectrum of the matrix **A**, and the columns of a unitary transforming matrix **T**, for which $\mathbf{B} = \mathbf{T}^{-1}\mathbf{AT}$, consist of the coordinates of an ortho-normal system of the n characteristic vectors of the matrix **A**, whose order corresponds to the order of the numbers λ_i in the diagonal of the matrix **B**. A similar assertion is true for a real symmetric matrix **A** if we replace unitary similarity by orthogonal similarity, where to obtain the orthogonal transforming matrix **T** it is necessary to take an ortho-normal system of n characteristic vectors with real coordinates.

EXAMPLE. For the matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{pmatrix}$$

to find the diagonal matrix B similar to it and the orthogonal matrix **T** for which $\mathbf{B} = \mathbf{T}^{-1}\mathbf{AT}$.

Because **A** is a real symmetric matrix, the problem is soluble. We will find the characteristic roots of the matrix **A**:

$$|\mathbf{A} - \lambda\mathbf{E}| = \begin{vmatrix} -\lambda & 1 & 1 & -1 \\ 1 & -\lambda & -1 & 1 \\ 1 & -1 & -\lambda & 1 \\ -1 & 1 & 1 & -\lambda \end{vmatrix}$$

(we add all the remaining rows to the first row, subtract the first column from the remaining ones and expand the determinant by the first row)

$$= (1-\lambda)\begin{vmatrix} -\lambda-1 & -2 & 0 \\ -2 & -\lambda-1 & 0 \\ 2 & 2 & -\lambda+1 \end{vmatrix}$$

$$= (1-\lambda)^2\begin{vmatrix} \lambda+2 & 2 \\ 2 & \lambda+1 \end{vmatrix} = (\lambda-1)^2(\lambda^2+2\lambda-3);$$

$$\lambda_1 = \lambda_2 = \lambda_3 = 1, \quad \lambda_4 = -3.$$

For the characteristic vectors with characteristic root 1 we obtain a system of linear equations equivalent to the one equation

$$x_1 - x_2 - x_3 + x_4 = 0.$$

It is necessary to find three solutions which form an ortho-normal system. We can take any solution as the first solution, for instance $\mathbf{a}_1 = (1, 1, 1, 1)$. We find the second solution by adding to the previous equation the condition of orthogonality to $\mathbf{a}_1$:

$$x_1 + x_2 + x_3 + x_4 = 0.$$

We can take, for instance, $\mathbf{a}_2 = (1, 1, -1, -1)$. To find the third solution we impose the further condition of orthogonality to $\mathbf{a}_2$:

$$x_1 + x_2 - x_3 - x_4 = 0,$$

which determines the solution to within a numerical factor. We take $\mathbf{a}_3 = (1, -1, 1, -1)$. The characteristic vector with characteristic root $\lambda_4 = -3$ can be found from the system of equations

$$\begin{aligned} 3x_1 + x_2 + x_3 - x_4 &= 0, \\ x_1 + 3x_2 - x_3 - x_4 &= 0, \\ x_1 - x_2 + 3x_3 + x_4 &= 0, \\ -x_1 + x_2 + x_3 + 3x_4 &= 0. \end{aligned}$$

Adding all the equations together and dividing by 4, we find that $x_1 + x_2 + x_3 + x_4 = 0$. Subtracting this equation from the first and second equations of the preceding system, we find $x_1 = x_4$, $x_2 = x_3$, whence $x_1 + x_2 = x_3 + x_4 = 0$. We can take $\mathbf{a}_4 = (1, -1, -1, 1)$. Norming the vectors and writing their coordinates in columns, we obtain the orthogonal matrix

$$\mathbf{T} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

which transforms the matrix $\mathbf{A}$ to the diagonal form $\mathbf{B} = \{1, 1, 1, -3\}$.

A matrix $\mathbf{A}$ is skew-Hermitian if and only if it is unitarily similar to a diagonal matrix $\mathbf{B}$ with purely imaginary elements along its diagonal (some of which may be equal to zero). A real matrix $\mathbf{A}$ is skew-symmetric if and only if it is orthogonally similar to a canonical skew-symmetric matrix

$$\mathbf{B} = \left\{ \begin{pmatrix} 0 & -a_1 \\ a_1 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & -a_q \\ a_q & 0 \end{pmatrix}, 0, \ldots, 0 \right\}. \tag{2.59}$$

Cells of the first and second order may not occur. Here the cells of the first order correspond to the zero characteristic roots of the matrix $\mathbf{A}$, and of the second order to pairs of conjugate characteristic numbers $\pm ia_k$ with $a_k > 0$, $k = 1, 2, \ldots, q$ ([6], p. 232–3). We can construct an orthogonal

system of characteristic vectors $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n$, where $\mathbf{z}_{2k-1} = \mathbf{x}_k + i\mathbf{y}_k$ corresponds to the value ia_k, and $\mathbf{z}_{2k} = \mathbf{x}_k - i\mathbf{y}_k$ to the value $-ia_k$, where the coordinates of the vectors $\mathbf{x}_k$, $\mathbf{y}_k$ are real, $k = 1, 2, \ldots, q$. The vectors $\mathbf{x}_k, \mathbf{y}_k$ will be orthogonal and of the same length ([20], Problem 1610). We can normalize them. For zero characteristic numbers we can also take the vectors $\mathbf{z}_{2q+1}, \ldots, \mathbf{z}_n$ as real and ortho-normal. By writing the coordinates of the vectors in the order

$$\mathbf{y}_1, \mathbf{x}_1, \mathbf{y}_2, \mathbf{x}_2, \ldots, \mathbf{y}_q, \mathbf{x}_q, \mathbf{z}_{2q+1}, \ldots, \mathbf{z}_n$$

in columns, we obtain an orthogonal matrix $\mathbf{T}$ which transforms the matrix $\mathbf{A}$ to the canonical form (2.59).

Matrices with positive spectra. The Hermitian (in particular, real symmetric) matrix $\mathbf{A}$ is called *a matrix with positive* (or *non-negative*) *spectrum*, if all its characteristic roots are positive (or, non-negative). In order that the Hermitian matrix $\mathbf{A}$ of order n should be a matrix with positive spectrum, it is necessary and sufficient that all its angular minors should be positive, i.e.

$$A\begin{pmatrix} 1 & 2 \ldots k \\ 1 & 2 \ldots k \end{pmatrix} > 0, \quad k = 1, 2, \ldots, n. \tag{2.60}$$

In order that the Hermitian matrix $\mathbf{A}$ of order n should be a matrix with non-negative spectrum, it is necessary and sufficient that all its principal (and not only its angular) minors should be non-negative. Thus, for instance, all the angular minors of the matrix $\begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$ are non-negative, but its spectrum $\lambda_1 = 0$, $\lambda_2 = -1$ is not non-negative.

For any matrix $\mathbf{H}$ with non-negative spectrum there exists a unique matrix $\mathbf{F}$, also with non-negative spectrum, such that $\mathbf{F}^2 = \mathbf{H}$. This matrix $\mathbf{F}$ is called the *arithmetical root* of the matrix $\mathbf{H}$ and is denoted by $\sqrt{\mathbf{H}}$. For any matrix $\mathbf{A}$ both the matrices $\mathbf{AA}^*$ and $\mathbf{A}^*\mathbf{A}$ are matrices with non-negative spectrum, and for a non-singular matrix $\mathbf{A}$ they are matrices with positive spectra. The matrix $\mathbf{AA}^*$ is called the *left*, and the matrix $\mathbf{A}^*\mathbf{A}$ the *right norm* of the matrix $\mathbf{A}$. The matrix $\sqrt{(\mathbf{AA}^*)}$ is called the *left* and the matrix $\sqrt{(\mathbf{A}^*\mathbf{A})}$ the *right modulus* of the matrix $\mathbf{A}$. For any matrix $\mathbf{H}$ of order n with non-negative spectrum the matrix $\sqrt{\mathbf{H}}$ can be represented in the form of a polynomial of degree not higher than the $(n-1)$th in the matrix $\mathbf{H}$ with real coefficients ([6], p. 225).

The polar decomposition of matrices. Any square matrix **A** with complex elements can be represented in each of the following two ways:

$$\mathbf{A} = \mathbf{HU}, \quad \mathbf{A} = \mathbf{U}_1\mathbf{H}_1, \tag{2.61}$$

where **H** and $\mathbf{H}_1$ are Hermitian matrices with non-negative spectra and **U** and $\mathbf{U}_1$ are unitary matrices. Here the matrices **H** and $\mathbf{H}_1$ are determined uniquely by the matrix **A**. The matrix **H** is the left, and the matrix $\mathbf{H}_1$ the right modulus of the matrix **A**. If the matrix **A** is non-singular, then the matrices **U** and $\mathbf{U}_1$ are also determined uniquely, and the matrices **H** and $\mathbf{H}_1$ will be matrices with positive spectra. If the matrix **A** has real elements, then **H** and $\mathbf{H}_1$ are real symmetric with non-negative spectra, and **U** and $\mathbf{U}_1$ will be orthogonal matrices. The representation (2.61) is the matrix analogue of the trigonometrical form of the complex number $z = r(\cos\varphi + i\sin\varphi)$ ([6], p. 225, Theorem 8; p. 233, Theorem 9). There is also a matrix analogue of the representation of a complex number in the form $z = re^{i\varphi}$ (see [6], p. 227, (87), (88)).

If **A** and **B** are Hermitian matrices, where **A** is a matrix with positive spectrum, then the matrix **AB** has real spectrum. If the matrices **A** and **B** are matrices with non-negative spectra and at least one of them is non-singular, then the matrix **AB** has real non-negative spectrum ([8], p. 138, or [20], Problems 1602, 1603). The sum of a finite number of matrices with non-negative spectra will be a matrix of the same type, and if at least one of the summands is a matrix with positive spectrum, then the sum will be the same ([20], Problem 1604). A matrix of rank r with non-negative spectrum can be represented in the form of a sum of r (but not less than r) matrices of rank 1 with non-negative spectra ([20], Problem 1605). A matrix **A** of order n is a matrix with non-negative spectrum of rank 1 if and only if it can be represented in the form $\mathbf{A} = \mathbf{X}^*\mathbf{X}$, where **X** is a non-zero row of n numbers ([20], Problem 1606).

If the matrices $\mathbf{A} = (a_{ij})_1^n$, $\mathbf{B} = (b_{ij})_1^n$ are matrices with non-negative spectra, then the matrix $\mathbf{C} = (a_{ij}b_{ij})_1^n$ obtained by multiplying together corresponding elements of the matrices **A** and **B**, will be of the same type ([20], Problem 1607).

Cayley's formulae. These formulae establish a connexion between Hermitian (or skew-Hermitian) and unitary matrices. If **F** is any Hermitian matrix, then the matrix

$$\mathbf{U} = (\mathbf{E} + i\mathbf{F})(\mathbf{E} - i\mathbf{F})^{-1} = (\mathbf{F} - i\mathbf{E})(\mathbf{F} + i\mathbf{E})^{-1} \tag{2.62}$$

will be unitary, having no characteristic root equal to -1. The second expression for $\mathbf{U}$ can be obtained from the first by removing a factor i from the first set of brackets, and $-i$ from the second set of brackets. If $\mathbf{U}$ is a unitary matrix which has no characteristic root equal to -1, then the matrix

$$\mathbf{F} = i(\mathbf{E} - \mathbf{U})(\mathbf{E} + \mathbf{U})^{-1} \tag{2.63}$$

will be Hermitian. The formulae (2.62) and (2.63) are inverses of each other and establish a one–one correspondence between all Hermitian matrices and all unitary matrices which have no characteristic root equal to -1 ([6], p. 228, (95)).

If $\mathbf{F}$ is any Hermitian matrix, then $\mathbf{K} = -i\mathbf{F}$ is a skew-Hermitian matrix, therefore formulae (2.62) and (2.63) can be combined in one formula:

$$\mathbf{U} = (\mathbf{E} - \mathbf{K})(\mathbf{E} + \mathbf{K})^{-1} \tag{2.64}$$

([6], p. 235, (123), (124)). If one of the matrices $\mathbf{U}$, $\mathbf{K}$ in equation (2.64) is skew-Hermitian (in particular, real skew-symmetric), then the other will be unitary (respectively, orthogonal), having no characteristic root equal to -1, and conversely. Formula (2.64) establishes a one–one correspondence between all skew-Hermitian (or real skew-symmetric) matrices and all unitary (respectively, orthogonal) matrices which have no characteristic root equal to -1. Here all non-singular skew-Hermitian (real skew-symmetric) matrices are mapped onto all the unitary (orthogonal) matrices which do not have characteristic roots equal to ± 1, and all the singular skew-Hermitian (real skew-symmetric) matrices are mapped onto all the unitary (orthogonal) matrices which have a characteristic root equal to $+1$, but do not have a characteristic root equal to -1 ([20], Problems 1614, 1615).

18. Normal matrices.

A matrix $\mathbf{A}$ with complex elements is called *normal* if it commutes with its conjugate matrix $\mathbf{A}^* = \overline{\mathbf{A}}'$, i.e. if $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$. In particular, a real matrix $\mathbf{A}$ is normal if $\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A}$. It is easy to verify that unitary, Hermitian, skew-Hermitian and real orthogonal, symmetric, and skew-symmetric matrices are special cases of normal matrices. A normal matrix $\mathbf{A}$ is unitary, Hermitian or skew-Hermitian if and only if its characteristic roots are equal in modulus to one, real numbers or purely imaginary numbers respectively (including cases when they are equal to zero).

A matrix **A** is normal if and only if in the representation (2.57) (p. 110) the Hermitian matrices **B** and **C** commute or in the representation (2.61) (p. 114) the matrices **H**, **U** and also $\mathbf{H}_1$, $\mathbf{U}_1$ commute ([6], p. 219, Theorem 3, and p. 225, Theorem 8). A matrix **A** is normal if and only if it is unitarily similar to a diagonal matrix **B** ([6], p. 222, Theorem 4). Here the columns of the unitary matrix **T** for which $\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ consist of the coordinates of an ortho-normal system of characteristic vectors of the matrix **A**.

A matrix **A** of order n with real elements is normal if and only if it is orthogonally similar to a canonical normal matrix of the form

$$\mathbf{B} = \left\{ \begin{pmatrix} a_1 & -b_1 \\ b_1 & a_1 \end{pmatrix}, \ldots, \begin{pmatrix} a_q & -b_q \\ b_q & a_q \end{pmatrix}, c_{2q+1}, \ldots, c_n \right\}, \quad (2.65)$$

where all the numbers a_k, b_k, c_k are real and $b_k > 0$, $k = 1, 2, \ldots, q$. Cells of the first and second order may not occur ([6], p. 232, (107)). Cells of the first order correspond to the real roots c_k, $k = 2q+1, \ldots, n$, and cells of the second order to pairs of complex conjugate roots $a_k \pm ib_k$, $k = 1, 2, \ldots, q$, of the characteristic polynomial of the matrix **A**. The transforming orthogonal matrix **T** is constructed by the same method as in the case of an orthogonal matrix (pp. 105–6). From the canonical form (2.65) for a real normal matrix we obtain, in particular, the canonical forms:

(1) for an orthogonal matrix, if

$$a_k^2 + b_k^2 = 1, \quad k = 1, 2, \ldots, q; \quad c_k = \pm 1, \quad k = 2q+1, \ldots, n;$$

(2) for a real symmetric matrix, if $q = 0$;

(3) for a real skew-symmetric matrix, if

$$a_k = 0, \quad k = 1, 2, \ldots, q; \quad c_k = 0, \quad k = 2q+1, \ldots, n.$$

If the matrix **A** is normal, then each of the conjugate matrices **A** and **A*** can be represented in the form of a polynomial in the other matrix ([6], p. 222, 2°).

The characteristic vectors of a normal matrix corresponding to distinct characteristic roots are orthogonal ([20], Problem 1628).

A matrix **A** is normal if and only if the matrices **A** and **A*** have the same characteristic vectors ([20], Problem 1630). Here we are considering both real and complex characteristic roots and characteristic vectors. If x is a characteristic vector of the normal matrix **A** with value λ_0, then the same vector

x will be characteristic for the matrix **A*** with the conjugate value $\overline{\lambda}_0$ ([20], Problem 1627).

If two normal (or real normal) matrices **A** and **B** are similar, then they are also unitarily (respectively, orthogonally) similar ([6], p. 236, Theorem 10).

Any (finite or infinite) set of pairwise commuting normal matrices has a common ortho-normal basis of characteristic vectors, and this means that, by means of the same unitary matrix **T**, all the matrices of this set can be reduced to diagonal form ([6], p. 237, Theorem 11′). Here if all the matrices of the given set are real, then there exists an orthogonal matrix **T** which reduces each of the matrices of this set to its canonical form (2.65), where orthogonal matrices are reduced to the form (2.56), symmetric matrices to real diagonal form and skew-symmetric matrices to the form (2.59) ([6], p. 238, Theorem 12′).

2. Bilinear and Quadratic Forms

1. Equivalence. The diagonal and normal forms. The methods of Lagrange and Jacobi for reduction to diagonal form.

A *bilinear form* of two systems of n variables

$$x_1, x_2, \ldots, x_n \quad \text{and} \quad y_1, y_2, \ldots, y_n$$

is a polynomial with numerical coefficients which is simultaneously of the first degree with respect to the variables of each of these systems. If we introduce the vectors

$$\mathbf{x} = (x_1, x_2, \ldots, x_n),$$
$$\mathbf{y} = (y_1, y_2, \ldots, y_n),$$

then the bilinear form will be a function of these vectors. Therefore we make use of the notation $A(\mathbf{x}; \mathbf{y})$ for its:

$$A(\mathbf{x}; \mathbf{y}) = \sum_{i,j=1}^{n} a_{ij} x_i y_j. \tag{2.66}$$

The matrix $\mathbf{A} = (a_{ij})_1^n$ of the coefficients of a bilinear form is called the *matrix* of this *form*, and the rank of this matrix is called the *rank of the bilinear form*. If we set $y_i = x_i$, $i = 1, 2, \ldots, n$, then the bilinear form reduces to a homogeneous polynomial of the second degree in the n variables $x_1, x_2, \ldots, x_n$, called a *quadratic form* in these variables.

Thus, to every bilinear form there corresponds a unique quadratic form. The same quadratic form may correspond to two distinct bilinear forms. For instance, the quadratic form $x_1^2 + 2x_1x_2 - x_2^2$ corresponds to the bilinear forms $x_1y_1 + x_1y_2 + x_2y_1 - x_2y_2$ and $x_1y_1 + 3x_1y_2 - x_2y_1 - x_2y_2$.

To establish a one–one correspondence between bilinear and quadratic forms we have to single out special classes of bilinear forms.

A bilinear form is called *symmetric* if its matrix is symmetric (p. 109). A symmetric bilinear form is unchanged if its two systems of variables are interchanged, i.e. $A(\mathbf{x}; \mathbf{y}) = A(\mathbf{y}; \mathbf{x})$. Any quadratic form can be obtained from a unique symmetric bilinear form, which is called the *polar bilinear form* of the given quadratic form. The matrix of the polar bilinear form is called the *matrix* of the given *quadratic form*. Thus, the quadratic form can be written in the form

$$A(\mathbf{x}; \mathbf{x}) = \sum_{i,j=1}^{n} a_{ij}x_ix_j, \quad a_{ij} = a_{ji}, \quad i, \ j = 1, 2, \ldots, n. \quad (2.67)$$

The expression on the right of this equation can be taken as the definition of a quadratic form. The matrix of a quadratic form $\mathbf{A} = (a_{ij})_1^n$, is symmetric by definition.

The rank of a matrix $\mathbf{A}$ of a quadratic form is called the *rank* of this form, and the determinant $|\mathbf{A}|$ the *discriminant* of the quadratic form. A quadratic form is called *non-degenerate* if $|\mathbf{A}| \neq 0$, and *degenerate* (or *singular*) if $|\mathbf{A}| = 0$.

The polar bilinear form of a given quadratic form can be expressed in the following way in terms of this quadratic form:

$$A(\mathbf{x}; \mathbf{y}) = \frac{1}{2}[A(\mathbf{x} + \mathbf{y}; \mathbf{x} + \mathbf{y}) - A(\mathbf{x}; \mathbf{x}) - A(\mathbf{y}; \mathbf{y})] \quad (2.68)$$

([8], p. 53).

If the coefficients a_{ij} are real, then the bilinear and quadratic forms are called *real*, and if they are complex, they are called *complex*. For a real form the values of the variables x_i, y_i are usually presumed to be real, and for a complex one they are presumed to be complex. A series of properties is also retained for forms with their coefficients from any number field (Introduction) or any field of characteristic not equal to two (Chapter IV).

We can define the bilinear form $A(\mathbf{x}; \mathbf{y})$ as a function of two vectors $\mathbf{x}, \mathbf{y}$ of an axiomatically defined linear space, which is linear in each argument, and a quadratic form $A(\mathbf{x}; \mathbf{x})$

as a function of the vector x, obtained from the bilinear form on the substitution $\mathbf{y} = \mathbf{x}$ ([8], section 4). However, we will use the definition given above of a bilinear and a quadratic form as polynomials of the forms (2.66) and (2.67).

Let $\mathbf{X}$ and $\mathbf{Y}$ be columns of variables $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$. The transposed matrices $\mathbf{X}'$; $\mathbf{Y}'$ will be rows of the same variables. By means of the multiplication of matrices the bilinear form (2.66) and the quadratic form (2.67) can be written in the form

$$A(\mathbf{x}; \mathbf{y}) = \mathbf{X}'\mathbf{A}\mathbf{Y}, \tag{2.69}$$

$$A(\mathbf{x}; \mathbf{x}) = \mathbf{X}'\mathbf{A}\mathbf{X}. \tag{2.70}$$

If we apply a non-singular linear transformation of variables

$$x_i = \sum_{j=1}^{n} s_{ij} u_j;\ y_i = \sum_{j=1}^{n} t_{ij} v_j;\ i = 1, 2, \ldots, n;\ |s_{ij}|_1^n \neq 0 \neq |t_{ij}|_1^n,$$

to the bilinear form (2.66), or in the matrix form

$$X = SU,\ Y = TV, \tag{2.71}$$

then, substituting expression (2.71) in (2.69), we obtain a new bilinear form

$$B(\mathbf{u}; \mathbf{v}) = \mathbf{U}'\mathbf{B}\mathbf{V} = \sum_{i,j=1}^{n} B_{ij} u_i v_j,$$

whose matrix $\mathbf{B}$ is connected with the matrix $\mathbf{A}$ of the original form by the equation

$$\mathbf{B} = \mathbf{S}'\mathbf{A}\mathbf{T} \tag{2.72}$$

([14], p. 209). In particular, if we apply the same transformation to the variables x_i and y_i, then $\mathbf{S} = \mathbf{T}$ and (2.72) reduces to

$$\mathbf{B} = \mathbf{T}'\mathbf{A}\mathbf{T},\quad |\mathbf{T}| \neq 0. \tag{2.73}$$

From this equation it is easy to deduce that if the bilinear form $\mathbf{A}(\mathbf{x}; \mathbf{y})$ were symmetric, then the new form $\mathbf{B}(\mathbf{u}; \mathbf{v})$ will also be symmetric. Therefore if we apply the non-singular linear transformation of variables

$$x_i = \sum_{j=1}^{n} t_{ij} y_j,\quad i = 1, 2, \ldots, n,\quad |\mathbf{T}| = |t_{ij}|_1^n \neq 0, \tag{2.74}$$

to the quadratic form (2.67), then we obtain a new quadratic form

$$B(\mathbf{y};\ \mathbf{y}) = \sum_{i,j=1}^{n} b_{ij} y_i y_j, \tag{2.75}$$

whose matrix **B** is connected with the matrix **A** of the original form (2.67) by the equation (2.73) ([6], p. 239, (2) and (4)). From this it follows that under a non-singular linear transformation of the variables the discriminant of a quadratic form is multiplied by the square of the determinant of the transformation, $|\mathbf{B}| = |\mathbf{A}| \cdot |\mathbf{T}|^2$, in particular, under a real non-singular transformation of a real form the sign of its discriminant is unchanged.

Two quadratic forms obtained from each other by non-singular linear transformations of the variables are called *equivalent* (or *congruent*). The matrices **A** and **B**, connected by the equation (2.73), are also called *congruent*. If the coefficients of the quadratic forms and the coefficients of the linear transformation are real, then the forms are called *real equivalent*; if they are complex, they are called *complex equivalent*; if they belong to some number field K, they are *called equivalent over the field* K. The equivalence of quadratic forms possesses the following properties: (1) reflexitivity: every form is equivalent to itself; (2) symmetry: if the first form is equivalent to the second, then the second is equivalent to the first; (3) transitivity: if the first form is equivalent to a second, and the second to a third, then the first is equivalent to the third. From this it follows that all the quadratic forms with coefficients from some number field K in the same variables can be divided into non-intersecting classes of forms equivalent over K. In particular, we can consider the classes of real or complex equivalent forms.

From equation (2.73) it follows that equivalent quadratic forms have the same rank. In the case of any number field K equality of ranks is insufficient for the equivalence of quadratic forms.

A quadratic form is called diagonal if it does not contain any products of variables, i.e. if it has the form

$$C(\mathbf{y};\ \mathbf{y}) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2. \tag{2.76}$$

Some of the coefficients λ_i may be equal to zero. If the forms (2.67) and (2.76) are equivalent, then the form (2.76) is called the *diagonal form* of the form (2.67).

Any quadratic form with coefficients from any number field K (or any field of characteristic not equal to two) can be reduced to diagonal form over this field by means of a non-singular linear transformation of variables. We will mention two methods of finding such a transformation.

Lagrange's method. Suppose that we have to reduce to diagonal form the quadratic form (2.67). If the form is zero, i.e. all its coefficients a_{ij} are equal to zero, then it is already diagonal. If the form is not zero, but all the coefficients a_{ii} of the squares of variables are equal to zero, then there exists a coefficient $a_{ij} \neq 0$, $i \neq j$. The transformation

$$x_i = y_1, \quad x_j = y_2, \quad x_1 = y_i, \quad x_2 = y_j, \quad x_k = y_k, \quad k \neq i, j, 1, 2 \tag{2.77}$$

is non-singular and reduces the form to a form where $a_{12} \neq 0$. Thus, we can assume that $a_{12} \neq 0$ already in the original form. Then the transformation

$$x_1 = y_1 - y_2, \quad x_2 = y_1 + y_2, \quad x_k = y_k, \quad k = 3, 4, \ldots, n, \tag{2.78}$$

reduces the form to a form containing a square of the first variable. If the original form does not contain x_1^2, but contains x_i^2, then the transformation

$$x_1 = y_i, \quad x_i = y_1, \quad x_k = y_k, \quad k_k \neq 1, \ i \tag{2.79}$$

reduces the form to a form containing the square of the first variable. Thus without loss of generality we can assume that $a_{11} \neq 0$. Then, singling out all the terms which contain x_1, we obtain

$$\begin{aligned} A(\mathbf{x}; \mathbf{x}) &= a_{11}x_1^2 + 2a_{12}x_1x_2 + \ldots + 2a_{1n}x_1x_n + A_1(\mathbf{x}; \mathbf{x}) \\ &= \frac{1}{a_{11}}(a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n)^2 + A_2(\mathbf{x}; \mathbf{x}), \end{aligned}$$

where $A_2(\mathbf{x}; \mathbf{x}) = A_1(\mathbf{x}; \mathbf{x}) - (1/a_{11})(a_{12}x_2 + \ldots + a_{1n}x_n)^2$ is a form which does not contain the first variable. By applying the transformation

$$\left.\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n, \\ y_2 &= x_2, \\ &\ldots\ldots\ldots\ldots\ldots \\ y_n &= x_n, \end{aligned}\right\} \tag{2.80}$$

we single out one square from the given form, i.e. we reduce it to the form $(1/a_{11})y_1^2 + A_2(\mathbf{y}; \mathbf{y})$, where $A_2(\mathbf{y}, \mathbf{y})$ does not

contain y_1. We treat the form $A(\mathbf{y}; \mathbf{y})$ in precisely the same way, leaving the first variable unchanged (only changing the notation for it). Because all the transformations (2.77)–(2.80) are non-singular, after a finite number of steps we obtain a non-singular linear transformation which reduces the original form to diagonal form.

The number of squares with non-zero coefficients λ_i in the diagonal form of a quadratic form is equal to the rank of the original form and does not depend on the choice of the non-singular linear transformation which reduces the given form to diagonal form.

EXAMPLE 1. To find the diagonal form and a non-singular linear transformation of variables reducing it to this form, for the form

$$A(\mathbf{x}; \mathbf{x}) = x_1x_2 + x_1x_3 + x_2x_3.$$

Because there are no squares in this form, we perform the transformation

$$x_1 = y_1 - y_2, \quad x_2 = y_1 + y_2, \quad x_3 = y_3. \tag{2.81}$$

The form becomes

$$A_1(\mathbf{y}; \mathbf{y}) = y_1^2 - y_2^2 + 2y_1y_3 = (y_1 + y_3)^2 - y_2^2 - y_3^2.$$

Applying the transformation

$$z_1 = y_1 + y_3, \quad z_2 = y_2, \quad z_3 = y_3, \tag{2.82}$$

we obtain the diagonal form

$$A_2(\mathbf{z}; \mathbf{z}) = z_1^2 - z_2^2 - z_3^2.$$

From (2.82) we find

$$y_1 = z_1 - z_3, \quad y_2 = z_2, \quad y_3 = z_3.$$

Substituting these expressions in (2.81), we obtain the desired transformation

$$\begin{aligned} x_1 &= z_1 - z_2 - z_3, \\ x_2 &= z_1 + z_2 - z_3, \\ x_3 &= z_3. \end{aligned}$$

Jacobi's method. In order to apply this method the quadratic form (2.67) must satisfy certain special conditions. An exposition of Jacobi's method in geometrical form can be found in [8], section 6. We will give an exposition in algebraic form. Let the quadratic form (2.67) have rank r and all the angular minors, i.e. the minors which stand in the upper left-hand corner of the matrix of the form, from the first to the rth order be distinct from zero:

$$D_k = A\begin{pmatrix} 1 & 2 & \dots & k \\ 1 & 2 & \dots & k \end{pmatrix} \neq 0, \quad k = 1, 2, \dots, r. \tag{2.83}$$

Then the diagonal form of the form (2.67) can be found by means of Jacobi's formula

$$C(\mathbf{y};\ \mathbf{y}) = \frac{1}{D_1} y_1^2 + \frac{D_1}{D_2} y_2^2 + \ldots + \frac{D_{r-1}}{D_r} y_r^2 \qquad (2.84)$$

([6], p. 245, (26)). It is easier not to calculate the coefficients of this diagonal form immediately, but in another way, which also gives expressions for the new variables y_j in terms of the old ones x_i. In fact, by virtue of (2.83) $a_{11} \neq 0$, and, by subtracting the first row of the matrix **A**, multiplied by suitable factors, from the remaining ones, we can reduce this matrix to the form

$$\mathbf{A}_1 = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ 0 & a'_{22} & \ldots & a'_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ 0 & a'_{n2} & \ldots & a'_{nn} \end{pmatrix}, \quad a'_{22} \neq 0.$$

Now all the elements of the second column which lie below a'_{22} also vanish, and so on, until we reduce the matrix to a triangular form **G**, in which all the rows, beginning with the $(r+1)th$, consist of zeros alone. We replace the zeros in these rows by any elements such that we obtain a non-singular triangular matrix. For instance, we can place 1's in the principal diagonal of the last $n-r$ rows, and leave the remaining elements zero. We obtain a matrix of the form:

$$\mathbf{T} = \begin{pmatrix} t_{11} & t_{12} & \ldots & t_{1r} & t_{1,r+1} & \ldots & t_{1n} \\ 0 & t_{22} & \ldots & t_{2r} & t_{2,r+1} & \ldots & t_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & t_{rr} & t_{r,r+1} & \ldots & t_{rn} \\ 0 & 0 & \ldots & 0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 1 \end{pmatrix}.$$

This matrix **T** will also be the matrix of the coefficients in the expressions for y_j in terms of x_i. The transformation inverse to that which we have constructed reduces the original form (2.67) to the diagonal form (2.84). The elements of the matrices **T** and **A** are connected by the equations

$$t_{ij} = \frac{\mathbf{A}\begin{pmatrix} 1 & 2 & \ldots & i-1 & i \\ 1 & 2 & \ldots & i-1 & j \end{pmatrix}}{D_{i-1}};$$

$$j = i,\ i+1,\ \ldots,\ n; \quad i = 1,\ 2,\ \ldots,\ r; \quad D_0 = 1. \qquad (2.85)$$

However, in practice it is more convenient to compute the reduction of the matrix **A** to triangular form by the method given above. Here it is easy to determine the rank of the matrix **A** and to verify the condition (2.83) of applicability of Jacobi's method. From (2.85) we find

$$t_{ii} = \frac{D_i}{D_{i-1}}, \quad i = 1, 2, \ldots, r, \quad D_0 = 1. \tag{2.86}$$

Therefore the diagonal form (2.84) can be rewritten thus:

$$C(\mathbf{y}; \mathbf{y}) = \sum_{i=1}^{r} \frac{1}{t_{ii}} y_i^2, \tag{2.87}$$

which enables us to find it by means of the matrix **T**, without calculating the minors $\mathbf{D}_i$.

EXAMPLE 2. To find the diagonal form and the corresponding non-singular transformation of variables for the quadratic form

$$A(\mathbf{x}; \mathbf{x}) = x_1^2 + x_2^2 - 2x_3^2 + \frac{2}{3} x_4^2 - 4x_1x_2 + 2x_1x_3 - 2x_1x_4 + 2x_2x_3 + 2x_2x_4 .$$

We reduce the matrix of the form to triangular form, denoting the equivalence of matrices by means of the sign $\sim$:

$$\begin{pmatrix} 1 & -2 & 1 & -1 \\ -2 & 1 & 1 & 1 \\ 1 & 1 & -2 & 0 \\ -1 & 1 & 0 & \frac{2}{3} \end{pmatrix}$$

(we add the first row to the following ones, first multiplying it respectively by 2, -1, and 1)

$$\sim \begin{pmatrix} 1 & -2 & 1 & -1 \\ 0 & -3 & 3 & -1 \\ 0 & 3 & -3 & 1 \\ 0 & -1 & 1 & -\frac{1}{3} \end{pmatrix}$$

(we add the second row, multiplied by 1 and $-1/3$, to the third and fourth rows)

$$\sim \begin{pmatrix} 1 & -2 & 1 & -1 \\ 0 & -3 & 3 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

From (2.87) we find the diagonal form $C(\mathbf{y}\,;\mathbf{y}) = y_1^2 - \frac{1}{3}y_2^2$

$$y_1 = x_1 - 2x_2 + x_3 - x_4,$$
$$y_2 = -3x_2 + 3x_3 - x_4,$$
$$y_3 = x_3, \quad y_4 = x_4.$$

From this we find the desired transformation of variables

$$x_1 = y_1 - \frac{2}{3}y_2 + y_3 + \frac{1}{3}y_4,$$
$$x_2 = -\frac{1}{3}y_2 + y_3 - \frac{1}{3}y_4,$$
$$x_3 = y_3, \quad x_4 = y_4.$$

For another derivation of Jacobi's method, see [7], pp. 43–44.

When the conditions (2.83) are satisfied for a form of rank r, Lagrange's method can only be carried out by means of a transformation of the form (2.80) and gives the same result as Jacobi's method. However, both the diagonal form of the given quadratic form, and also the transformation which reduces it to the given canonical form, are *not defined uniquely*. Thus, if $a_{11} \neq 0$, by excluding the square of the first unknown we can remove from the bracket not only $1/a_{11}$, but any non-zero factor. This factor will also be the coefficient for the square of the first variable in the diagonal form. If, for instance, we remove the factor a_{11}, then instead of (2.80) we obtain the transformation

$$y_1 = x_1 + \frac{a_{12}}{a_{11}}x_2 + \ldots + \frac{a_{1n}}{a_{11}}x_n,$$
$$y_i = x_i \quad \text{when} \quad i > 1.$$

From the conditions (2.83) the form of rank r is reduced to a diagonal form (2.76) by means of several transformations of this form which give as a result the triangular transformation of the special form

$$\left.\begin{aligned} y_1 &= x_1 + c_{12}x_2 + \ldots + c_{1n}x_n, \\ y_2 &= x_2 + c_{23}x_3 + \ldots + c_{2n}x_n, \\ &\cdots\cdots\cdots\cdots \\ y_n &= x_n \end{aligned}\right\} \tag{2.88}$$

with determinant equal to 1. The inverse transformation will also be of the same special form

$$\left.\begin{aligned} x_1 &= y_1 + d_{12}y_2 + \ldots + d_{1n}y_n, \\ x_2 &= y_2 + d_{23}y_3 + \ldots + d_{2n}y_n, \\ &\cdots\cdots\cdots\cdots \\ x_n &= y_n. \end{aligned}\right\} \tag{2.89}$$

We can prove that under a transformation of the form (2.89) the angular minors of the matrix **A** of the quadratic form are unchanged, and therefore the diagonal form to which a form of rank r is reduced under the conditions (2.83) by a transformation of the form (2.89) must look like this:

$$C(\mathbf{y};\ \mathbf{y}) = D_1 y_1^2 + \frac{D_2}{D_1} y_2^2 + \ldots + \frac{D_r}{D_{r-1}} y_r^2 \qquad (2.90)$$

([20], Problems 1221, 1222).

Thus the diagonal form is not defined for a given class of equivalent quadratic forms. We indicate here the normal form for a complex form (and in the following section, for a real form), which is uniquely defined for every class of equivalent forms. Taking into account the change in the numbering of the variables y_i in the canonical form (2.76) we can assume that for a form of rank r the conditions

$$\lambda_i \neq 0, \quad \text{when} \quad i = 1, 2, \ldots, r;$$
$$\lambda_i = 0, \quad \text{when} \quad i = r+1, \ldots, n,$$

are satisfied. Then the transformation

$$z_i = \sqrt{\lambda_i}\, y_i \quad \text{when} \quad i = 1, 2, \ldots, r,$$
$$z_i = y_i \qquad \text{when} \quad i = r+1, \ldots, n$$

will be non-singular and reduces (2.76) to the form

$$D(\mathbf{z};\ \mathbf{z}) = z_1^2 + z_2^2 + \ldots + z_r^2. \qquad (2.91)$$

This form is called the *normal form* of the original form.

Two complex quadratic forms are complex equivalents if and only if they have the same normal form, i.e. if they have the same rank ([11], p. 175). Because the rank of a form of n variables can take any value from zero to n, the number of classes of complex equivalent quadratic forms in n variables is equal to $n+1$.

The canonical form of a bilinear form. Because the matrix of the quadratic form (2.67) and the matrix of its polar bilinear form under an identical transformation of variables are transformed by means of the same formula (2.73), then any symmetric bilinear form can be reduced to the canonical form

$$C(\mathbf{u};\ \mathbf{v}) = \lambda_1 u_1 v_1 + \lambda_2 u_2 v_2 + \ldots + \lambda_n u_n v_n \qquad (2.92)$$

by means of a non-singular linear transformation of variables (in fact, the transformation which reduces the correspond-

ing quadratic form to the diagonal form (2.76), applied to both series of variables x_i and y_i).

A non-symmetric bilinear form cannot be reduced to canonical form by means of an identical transformation of both systems of variables.

The bilinear form (2.66) is called *skew-symmetric* if its matrix **A** is skew-symmetric (p. 129). The rank of a skew-symmetric bilinear form is always an even number. Any skew-symmetric bilinear form with coefficients from some number field K can be reduced to the normal form

$$D(\mathbf{u}; \mathbf{v}) = u_1v_2 - u_2v_1 + u_3v_4 - u_4v_3 + \ldots + u_{r-1}v_r - u_rv_{r-1}, \quad (2.93)$$

where r is the rank of the bilinear form ([14], p. 217), by means of a non-singular linear transformation with coefficients from K, which is the same for both series of variables.

From the possibility of reducing the form to the forms (2.76), (2.91) and (2.93) there follow the corresponding properties of the matrices:

(1) Any symmetric matrix with elements from a number field K is congruent to a diagonal matrix over K.

(2) Any complex symmetric matrix is complex congruent to a diagonal matrix of the form $\{1, \ldots, 1, 0, \ldots, 0\}$, where the number of 1's is equal to the rank of the given matrix.

(3) Any skew-symmetric matrix of rank r with elements from a number field K is congruent over K to the matrix

$$\left\{\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, 0, \ldots, 0\right\},$$

where the number of cells of order 2 is $r/2$.

For the reduction of certain quadratic forms of special types to diagonal form, see [20], Problems 1193–8.

2. Real quadratic forms. The law of inertia. Indices of inertia and signature.

All the forms in this section are presupposed real, and by the equivalence of forms we shall mean their equivalence over the field of real numbers.

Lagrange's and Jacobi's methods of reducing a quadratic form to diagonal form, explained in the preceding section, do not go outside the domain of real numbers when they are applied to a real form. Therefore any real quadratic form is equivalent over the fields of real numbers to a diagonal form.

Taking into account the change in the numbering of the variables a form of rank r can be reduced to the form

$$C(\mathbf{y};\mathbf{y}) = \lambda_1 y_1^2 + \ldots + \lambda_p y_p^2 - \lambda_{p+1} y_{p+1}^2 - \ldots - \lambda_{p+q} y_{p+q}^2,$$
$$p+q=r, \quad \lambda_i > 0, \quad i = 1, 2, \ldots, r. \tag{2.94}$$

Here the numbers p and q may equal zero. The real non-singular transformation of variables

$$z_i = \sqrt{\lambda_i}\, y_i, \quad i = 1, 2, \ldots, r; \quad z_i = y_i, \quad i = r+1, \ldots, n,$$

reduces the diagonal form (2.94) to the form

$$D(\mathbf{z};\mathbf{z}) = z_1^2 + \ldots + z_p^2 - z_{p+1}^2 - \ldots - z_{p+q}^2,$$
$$p+q=r, \tag{2.95}$$

called the *normal form* of the real quadratic form.

The law of inertia for quadratic forms. The number p of positive and the number q of negative coefficients in the canonical form (2.94) or in the normal form (2.95) of a given quadratic form do not depend on the choice of the real non-singular transformation of variables which reduces the original form to this form ([11], p. 176). The numbers p and q are called, respectively, the *positive and negative indices of inertia*, and their difference $s = p - q$ is called the *signature* of the quadratic form. The indices of inertia, the rank and signature are connected in the following way:

$$\left.\begin{aligned} r &= p+q, & s &= p-q; \\ p &= \frac{1}{2}(r+s), & q &= \frac{1}{2}(r-s), \end{aligned}\right\} \tag{2.96}$$

therefore r and s determine p and q uniquely and conversely.

From the law of inertia it follows that two quadratic forms are equivalent (in the real domain) if and only if they have the same normal form, or the same indices of inertia, or the same rank and signature.

The number of classes of real equivalent quadratic forms in n variables is equal to $(n+1)(n+2)/2$. The number of such classes for given rank r is equal to $r+1$, and for given signature s is equal to $[(n-|s|)/2]+1$, where $[a]$ denotes the integral part of the number a ([20], Problems 1203, 1205).

From the expressions for the diagonal form in the form (2.84) or (2.90) of the preceding section there follows:

Jacobi's theorem. If for the quadratic form

$$A(\mathbf{x};\mathbf{x}) = \sum_{i,j=1}^{n} a_{ij}x_i x_j$$

of rank r the conditions (2.83) of the preceding section are satisfied, then the positive index of inertia p is equal to the number of unchanged signs, and the negative index of inertia q is equal to the number of changes of sign in the series of numbers

$$1, D_1, D_2, \ldots, D_r, \tag{2.97}$$

$$\left.\begin{aligned} p &= P(1, D_1, D_2, \ldots, D_r), \\ q &= V(1, D_1, D_2, \ldots, D_r). \end{aligned}\right\} \tag{2.98}$$

Therefore for the signature we have

$$s = r - 2q = r - 2V(1, D_1, D_2. \ldots, D_r) \tag{2.99}$$

([6], p. 246, Theorem 2).

Formula (2.99) for the signature also remains in force in certain cases when separate minors in the series (2.97) are equal to zero. In fact, we have:

Hildenfinger's rule. If for a quadratic form of rank r: (a) $D_r \neq 0$, (b) in the series of minors (2.97) there are not two zeros one after the other, then the signature of this form can be determined from equation (2.99), where in counting the number of changes in sign the zeros are omitted ([6], p. 246).

Frobenius's rule. If for a quadratic form of rank r: (a) $D_r \neq 0$, (b) in the series of minors (2.97) there are not three zeros one after the other, then the signature can be determined from the equation (2.99), where if $D_k = 0$, but $D_{k-1}D_{k+1} \neq 0$, then the zero is not counted, but if $D_k = D_{k+1} = 0$, but $D_{k-1}D_{k+2} \neq 0$, then

$$V(D_{k-1}, D_k, D_{k+1}, D_{k+2}) = \begin{cases} 1 & \text{when} \quad D_{k-1}D_{k+2} < 0, \\ 2 & \text{when} \quad D_{k-1}D_{k+2} > 0 \end{cases}$$

([6], p. 246).

If in the series of minors (2.97) there are three zeros, one after the other, then the signs of the minors D_k do not determine the signature of the form uniquely. Thus, for the form

$$2ax_1x_4 + bx_2^2 + cx_3^2$$

when $a \neq 0$, $bc > 0$ we have

$$D_1 = D_2 = D_3 = 0, \quad D_4 = -a^2bc < 0.$$

But when $b > 0$, $c > 0$ we will have $q = 1$, and when $b < 0$, $c < 0$ we will have $q = 3$. In the same way the signs of the minors (2.97) do not determine the signature of a form of rank r, if $D_1 \cdot D_2 \dots D_{r-1} \neq 0$, but $D_r = 0$. Thus, when $a \neq 0$, $a \neq b$ the form

$$ax_1^2 + ax_2^2 + bx_3^2 + 2ax_1x_2 + 2ax_1x_3 + 2ax_2x_3 = a(x_1 + x_2 + x_3)^2 + (b - a)x_3^2$$

has rank two, but $s = 0$ when $a > b > 0$, and $s = 2$ when $b > a > 0$, at least in the cases $D_1 = a > 0$, $D_2 = D_3 = 0$ ([6], p. 247).

A linear form is a homogeneous polynomial of the first degree, i.e. a polynomial of the form $a_1x_1 + a_2x_2 + \dots + a_nx_n$. The product of two linear forms in n variables is a quadratic form in these variables. However, not every quadratic form can be decomposed into the product of linear forms. In order that the quadratic form $A(\mathbf{x}, \mathbf{x}) = \sum_{i,j=1}^{n} a_{ij}x_ix_j$ with complex or real coefficients should decompose into the product of two linear forms, it is necessary and sufficient that the following conditions should be fulfilled: (1) for complex linear forms: the rank of the quadratic form $\leqslant 2$; (2) for real linear forms: the rank of the quadratic form $\leqslant 2$ and for rank 2 the signature is equal to zero ([11], p. 178). From the existence of the normal form (2.95) there follows this property for matrices: any real symmetric matrix is real congruent to a unique (for the given matrix) diagonal matrix of the form $\{1, \dots, 1, -1, \dots, -1, 0, \dots, 0\}$, where the number of elements ± 1 is equal to the rank of the given matrix (the elements $+1$, -1 or 0 may not appear).

3. Positive definite quadratic forms.

A real quadratic form

$$A(\mathbf{x}; \mathbf{x}) = \sum_{i,j=1}^{n} a_{ij}x_ix_j \tag{2.100}$$

is called *positive definite* if $A(\mathbf{x}; \mathbf{x}) > 0$ for any real values of the variables $x_1, x_2, \dots, x_n$, not all of which are equal to zero; *negative definite* if $A(\mathbf{x}; \mathbf{x}) < 0$ for such numbers; *non-negative* if $A(\mathbf{x}; \mathbf{x}) \geqslant 0$, and *non-positive* if $\mathbf{A}(\mathbf{x}; \mathbf{x}) \leqslant 0$ for any real

values of the variables. We will denote these four properties respectively by the inequalities

$$A(\mathbf{x};\mathbf{x})>0, \quad A(\mathbf{x};\mathbf{x})<0, \quad A(\mathbf{x};\mathbf{x})\geqslant 0, \quad A(\mathbf{x};\mathbf{x})\leqslant 0.$$

The forms $A(\mathbf{x};\mathbf{x})>0$ and $A(\mathbf{x};\mathbf{x})<0$ are called *definite*, and the forms $A(\mathbf{x};\mathbf{x})\geqslant 0$ and $A(\mathbf{x};\mathbf{x})\leqslant 0$ are called *semi-definite*. Definite forms are always non-degenerate. Forms which are semi-definite but not definite are always degenerate ([6], p. 248). If $A(\mathbf{x};\mathbf{x})>0$ (or $\geqslant 0$), then $-A(\mathbf{x};\mathbf{x})<0$ (respectively, $\leqslant 0$). Therefore every property of forms of one sign corresponds to a definite property of forms of the opposite sign.

Criteria for positive definite quadratic forms.

(1) In order that a real quadratic form in n unknowns should be positive definite, it is necessary and sufficient that it should reduce to a normal form consisting of n squares with coefficients $+1$, i.e. to the form

$$D(\mathbf{z};\mathbf{z})=z_1^2+z_2^2+\ldots+z_n^2, \tag{2.101}$$

or that one of the equations $p=n$, $s=n$ should be satisfied.

(2) *Sylvester's criterion.* In order that a real quadratic form (2.100) should be positive definite it is necessary and sufficient that all the angular minors of its matrix $\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_n$, where

$$D_k=\left|A\begin{pmatrix}1 & 2 & \ldots & k\\ 1 & 2 & \ldots & k\end{pmatrix}\right|, \quad k=1,2,\ldots,n, \tag{2.102}$$

should be positive

$$D_k>0, \quad k=1,2,\ldots,n \tag{2.103}$$

([11], p. 181).

From this we can deduce that if $A(\mathbf{x};\mathbf{x})>0$, then all the principal minors $A\begin{pmatrix}i_1 & i_2 & \ldots & i_k\\ i_1 & i_2 & \ldots & i_k\end{pmatrix}$ of its matrix are positive ([6], p. 249). In particular, all the coefficients of the squares of variables are positive: $a_{kk}>0$ $(k=1,2,\ldots,n)$.

(3) In order that the real form (2.100) should be non-negative, it is necessary and sufficient that all its principal minors should be non-negative ([6], p. 249, Theorem 4).

If some of the angular minors are non-negative, it is not sufficient to ensure that the form is non-negative.

(4) In order that a real quadratic form should be positive definite (non-negative), it is necessary and sufficient that all

the characteristic roots of its matrix should be positive (non-negative).

(5) In order that the real quadratic form (2.100) should be positive definite, it is necessary and sufficient that its matrix **A** should be able to be represented in the form

$$\mathbf{A} = \mathbf{C}'\mathbf{C}, \tag{2.104}$$

where **C** is a non-singular real matrix of order n; $\mathbf{C}'$ is the transposed matrix.

(6) In order that the real quadratic form (2.100) of rank r should be non-negative, it is necessary and sufficient that its matrix **A** have a representation of the form (2.104), where **C** is a real matrix of order n and rank r. We can assume that the first r rows of the matrix **C** are linearly independent, and the remaining rows are zero ([20], Problem 1210d).

If we go over to the form $-A(\mathbf{x}; \mathbf{x})$, from the properties (1)–(6) we can obtain the corresponding properties for negative definite and non-positive forms. We mention only the property corresponding to (2).

(7) In order that the real quadratic form (2.100) should be negative definite, it is necessary and sufficient that the signs of the angular minors (2.102) should alternate, and that $\mathbf{D}_1 < 0$.

(8) If both the quadratic forms

$$A(\mathbf{x}; \mathbf{x}) = \sum_{i,j=1}^{n} a_{ij} x_i x_j \text{ and } B(\mathbf{x}; \mathbf{x}) = \sum_{i,j=1}^{n} b_{ji} x_i x_j$$

are positive definite (or non-negative), then the form $\mathbf{C}(x; x) = \sum_{i,j=1}^{n} a_{ij} b_{ij} x_i x_j$ is also positive definite (non-negative) ([20], Problem 1220).

4. Orthogonal transformations of quadratic forms. The reduction of a quadratic form to principal axes.

A linear transformation of variables

$$x_i = \sum_{j=1}^{n} t_{ij} y_j, \quad i = 1, 2, \ldots, n, \tag{2.105}$$

is called *orthogonal* if its matrix $\mathbf{T} = (t_{ij})_1^n$ is orthogonal, i.e. $\mathbf{T}' = \mathbf{T}^{-1}$ (p. 103).

If we apply the orthogonal transformation of variables (2.105) to the quadratic from

$$A(\mathbf{x};\mathbf{x}) = \sum_{i,j=1}^{n} a_{ij}x_i x_j \tag{2.106}$$

then the equation

$$\mathbf{B} = \mathbf{T}'\mathbf{A}\mathbf{T}, \tag{2.107}$$

connecting the matrix **A** of the form (2.106) and the matrix **B** of the form obtained as a result of the transformation, can be rewritten thus:

$$\mathbf{B} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}. \tag{2.108}$$

In particular, if $A(\mathbf{x};\mathbf{x}) = x_1^2 + x_2^2 + \ldots + x_n^2$, i.e. $\mathbf{A} = \mathbf{E}$, then $\mathbf{B} = \mathbf{E}$ also, and $x_1^2 + x_2^2 + \ldots + x_n^2 \equiv y_1^2 + y_2^2 + \ldots + y_n^2$. This identity is necessary and sufficient for the transformation (2.105) to be orthogonal.

Thus the relation of congruence (2.107) becomes the relation of similarity (2.108). From the fact that any real symmetric matrix is orthogonally similar to a real diagonal matrix (p. 111) it follows that for any real quadratic form (2.106) there exists an orthogonal transformation of variables (2.105) which reduces it to the form

$$B(\mathbf{y};\mathbf{y}) = \sum_{i=1}^{n} \lambda_i y_i^2. \tag{2.109}$$

Under any such transformation the coefficients λ_i of the squares of variables turn out to be the characteristic roots (taking into account their multiplicity) of the matrix **A** of the given form ([6], p. 251, Theorem 7). Therefore the diagonal form (2.109) is defined uniquely to within the order of the coefficients λ_i.

If the form (2.106) has rank r, then we can assume that $\lambda_i \neq 0$ when $i = 1, 2, \ldots, r$ and $\lambda_i = 0$ when $i = r+1, \ldots, n$. The orthogonal transformation (2.105) which reduces the form 2.106) to the diagonal form (2.109) is not, in the general case, defined uniquely.

In order that the orthogonal transformation of variables (2.105) should reduce the real quadratic form (2.106) to the diagonal form (2.109), it is necessary and sufficient that the columns of the matrix **T** should form an ortho-normal system of characteristic vectors (with real coordinates) of the matrix **A** of the given form. The order of the columns in the matrix **T** must correspond to the order of the coefficients λ_i in the

canonical form (2.109). The columns which correspond to simple roots are defined uniquely to within sign, and those corresponding to multiple roots are not defined uniquely. The number of columns corresponding to a given root is equal to the multiplicity of this root.

EXAMPLE 1. To find an orthogonal transformation of variables which reduces to diagonal form the quadratic form

$$A(\mathbf{x};\mathbf{x}) = 2x_1x_2 + 2x_1x_3 - 2x_1x_4 - 2x_2x_3 + 2x_2x_4 + 2x_3x_4.$$

The matrix of this form

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{pmatrix}$$

was considered in Example 1 on p. 111. There we found the characteristic roots $\lambda_1 = \lambda_2 = \lambda_3 = 1$, $\lambda_4 = -3$ and the orthogonal matrix in whose columns stand the coordinates of the characteristic vectors is

$$\mathbf{T} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Therefore the orthogonal transformation of variables

$$x_1 = \frac{1}{2}(y_1 + y_2 + y_3 + y_4),$$

$$x_2 = \frac{1}{2}(y_1 + y_2 - y_3 - y_4),$$

$$x_3 = \frac{1}{2}(y_1 - y_2 + y_3 - y_4),$$

$$x_4 = \frac{1}{2}(y_1 - y_2 - y_3 + y_4)$$

reduces the given form to the form

$$B(\mathbf{y};\mathbf{y}) = y_1^2 + y_2^2 + y_3^2 - 3y_4^2.$$

The rank of the quadratic form (2.106) is equal to the number of characteristic roots of its matrix A distinct from zero, the indices of inertia to the number of positive and the number of negative characteristic roots, and the signature to the difference between these values. If under a continuous variation in the coefficients of a real form its rank remains unchanged, then its signature also remains unchanged, and conversely ([6], p. 251).

Remark. The reduction of a quadratic form to diagonal form by means of an orthogonal transformation of variables is also known as reducing this form to *principal axes*. This terminology is explained by the connexion with geometry. Suppose that we are given a central hypersurface of the second order in n-dimensional Euclidean space:

$$\sum_{i,j=1}^{n} a_{ij}x_ix_j = c, \quad a_{ij} = a_{ji}, \quad c \neq 0. \tag{2.110}$$

If we reduce the quadratic form on the left-hand side of the equation to canonical form by means of an orthogonal transformation, we obtain the equation

$$\sum_{i=1}^{n} \lambda_i y_i^2 = c, \tag{2.111}$$

which shows that the new axes of coordinates are the axes of symmetry, or, as they are called, the *principal axes* of the hypersurface. The directions of these axes are given by the vectors whose coordinates form the columns of the matrix $\mathbf{T}$ of the transformation (2.105). Taking into account the change in sign in one of the columns we can assume that $|\mathbf{T}| = 1$. Then the new system of coordinates is obtained from the old by means of some "rotation".

From the properties of commuting normal (in particular, real symmetric) matrices (p. 117) it follows that a set of two or more real quadratic forms has an orthogonal transformation of variables, simultaneously reducing to diagonal form each of the forms of the given set, if and only if the matrices of the forms of this set commute in pairs.

In geometrical form this means that two or more central surfaces of the second order have parallel principal axes if and only if their matrices commute in pairs.

The canonical form of a non-homogeneous polynomial of the second degree. By analogy with geometry we can consider the more general algebraic problem — the reduction of any (not necessarily homogeneous) polynomial of the second degree in n variables to canonical form by means of an orthogonal (not necessarily homogeneous) transformation of variables. Suppose that we are given the polynomial

$$\sum_{i,j=1}^{n} a_{ij}x_ix_j + 2\sum_{i=1}^{n} b_ix_i + c. \tag{2.112}$$

In the first place, we reduce the quadratic form of all the terms of the second degree to the diagonal form $\sum_{i=1}^{r} \lambda_i y_i^2$, where

r is the rank of this form, by means of an orthogonal transformation of variables

$$x_i = \sum_{j=1}^{n} p_{ij} y_j, \quad i = 1, 2, \ldots, n. \tag{2.113}$$

After this transformation the polynomial (2.112) takes the form

$$\left.\begin{aligned} &\sum_{i=1}^{r} \lambda_i y_i^2 + 2\sum_{i=1}^{n} b_i' y_i + c, \\ &\lambda_i \neq 0, \quad i = 1, 2, \ldots, r. \end{aligned}\right\} \tag{2.114}$$

Secondly, from the equation

$$\lambda_i y_i^2 + 2b_i' y_i = \lambda_i \left(y_i + \frac{b_i'}{\lambda_i}\right)^2 - \frac{b_i'^2}{\lambda_i}$$

it is clear that the parallel displacement

$$\left.\begin{aligned} y_i &= z_i - \frac{b_i'}{\lambda_i} \text{ when } i = 1, 2, \ldots, r, \\ y_i &= z_i \qquad\quad \text{when } i = r+1, \ldots, n \end{aligned}\right\} \tag{2.115}$$

reduces the polynomial (2.114) to the form

$$\sum_{i=1}^{r} \lambda_i z_i^2 + 2\sum_{i=r+1}^{n} b_i' z_i + c'; \tag{2.116}$$

thirdly, if not all the $b_{r+1}', \ldots, b_n'$ are equal to zero, then the $(n-r)$-dimensional vector

$$\frac{1}{\sqrt{\left(\sum_{i=r+1}^{n} b_i'^2\right)}} (b_{r+1}', \ldots, b_n') = (q_{r+1,\,r+1}, \ldots, q_{r+1,\,n})$$

can be supplemented to an ortho-normal basis for $(n-r)$-dimensional space by means of the vectors

$$\begin{aligned} &(q_{r+2,\,r+1}, \cdots, q_{r+2,\,n}), \\ &\cdots\cdots\cdots\cdots \\ &(q_{n,\,r+1}, \cdots, q_{n,\,n}). \end{aligned}$$

The transformation

$$u_i = z_i, \qquad \text{when } i = 1, 2, \ldots, r,$$

$$u_i = \sum_{j=r+1}^{n} q_{ij} z_j, \quad \text{when } i = r+1, \ldots, n$$

will be orthogonal. The inverse transformation

$$\left.\begin{aligned} z_i &= u_i, && \text{when } i = 1, 2, \ldots, r, \\ z_i &= \sum_{j=r+1}^{n} q_{ji} u_j && \text{when } i = r+1, \ldots, n \end{aligned}\right\} \tag{2.117}$$

(whose matrix is obtained by virtue of the orthogonality of the transposed matrix for the matrix of the preceding transformation) reduces the polynomial (2.116) to the form

$$\sum_{i=1}^{r} \lambda_i u_i^2 + 2\lambda_{r+1} u_{r+1} + c', \tag{2.118}$$

where

$$\lambda_{r+1} = \sqrt{\left(\sum_{i=r+1}^{n} b_i'^2\right)} \neq 0.$$

Finally, the transformation

$$u_i = v_i, \quad \text{when } i \neq r+1, \quad u_{r+1} = v_{r+1} - \frac{c'}{2\lambda_{r+1}} \tag{2.119}$$

reduces the polynomial (2.118) to the form

$$\sum_{i=1}^{r} \lambda_i v_i^2 + 2\lambda_{r+1} v_{r+1}. \tag{2.120}$$

The successive application of transformations (2.113), (2.115), (2.117) and (2.119) gives an orthogonal transformation (in the general case, not homogeneous) which reduces the original polynomial (2.112) to one of three canonical forms: either (2.120) ($\lambda_{r+1} \neq 0$), or

$$\sum_{i=1}^{r} \lambda_i v_i^2 \quad (c' = \lambda_{r+1} = 0), \tag{2.121}$$

or

$$\sum_{i=1}^{r} \lambda_i v_i^2 + a \quad (c' \neq 0, \lambda_{r+1} = 0). \tag{2.122}$$

If we consider the polynomial (2.112) as the left-hand side of the equation of a hypersurface of the second order in n-dimensional Euclidean space, then in the case of the canonical form (2.120) the surface has axes of symmetry, and in the cases (2.121) and (2.122) a centre of symmetry.

EXAMPLE 2. To find an orthogonal transformation which reduces to canonical form the polynomial

$$2x_1x_2 + 2x_1x_3 - 2x_1x_4 - 2x_2x_3 + 2x_2x_4 + 2x_3x_4 + 3x_1 + 3x_2 + x_3 + x_4 + x_5 + x_6 + 1.$$

The quadratic form of this polynomial was considered in Example 1. Because now the polynomial depends on six variables, the transformation (2.113), taking into account the result of Example 1, assumes the form

$$\left.\begin{aligned} x_1 &= \frac{1}{2}(y_1 + y_2 + y_3 + y_4), \quad x_2 = \frac{1}{2}(y_1 + y_2 - y_3 - y_4), \\ x_3 &= \frac{1}{2}(y_1 - y_2 + y_3 - y_4), \quad x_4 = \frac{1}{2}(y_1 - y_2 - y_3 + y_4), \\ x_5 &= y_5, \qquad\qquad x_6 = y_6. \end{aligned}\right\} \tag{2.123}$$

This transformation reduces the polynomial to the form

$$y_1^2 + y_2^2 + y_3^2 - 3y_4^2 + 4y_1 + 2y_2 + y_5 + y_6 + 1. \tag{2.124}$$

We set

$$z_1 = y_1 + 2, \quad z_2 = y_2 + 1, \quad z_i = y_i, \quad \text{when} \quad i = 3, 4, 5, 6.$$

From this we find a transformation analogous to (2.115)

$$y_1 = z_1 - 2, \quad y_2 = z_2 - 1, \quad y_i = z_i, \quad \text{when} \quad i = 3, 4, 5, 6. \tag{2.125}$$

This transformation reduces the polynomial (2.124) to the form

$$z_1^2 + z_2^2 + z_3^2 - 3z_4^2 + z_5 + z_6 - 4. \tag{2.126}$$

In addition, it is necessary to norm the vector $\mathbf{a}_1 = (1, 1)$ of the coefficients of z_5, z_6, and to supplement it to an ortho-normal system. It is more convenient, however, to construct an orthogonal system first, and then to norm the vectors. In the given example $n - r = 6 - 4 = 2$, and for the vector $\mathbf{a}_1$ it is only necessary to find one vector orthogonal to it. This vector we can find from the condition for orthogonality: $x_1 + x_2 = 0$. We can take $\mathbf{a}_2 = (1, -1)$. Norming, we obtain the matrix

$$\frac{1}{2}\sqrt{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

The inverse matrix coincides with the matrix itself by virtue of its orthogonality and symmetry. Therefore the transformation analogous to (2.117) has the form

$$\left.\begin{aligned} z_i &= u_i, \qquad \text{when} \qquad i = 1, 2, 3, 4 \\ z_5 &= \frac{1}{2}\sqrt{2}(u_5 + u_6), \quad z_6 = \frac{1}{2}\sqrt{2}(u_5 - u_6). \end{aligned}\right\} \tag{2.127}$$

The polynomial (2.126) takes the form

$$u_1^2 + u_2^2 + u_3^2 - 3u_4^2 + \sqrt{(2)}\,u_5 - 4. \tag{2.128}$$

Finally, setting $\sqrt{2}u_5 - 4 = \sqrt{2}(u_5 - 2\sqrt{2}) = \sqrt{2}v_5$, we find a transformation analogous to (2.119),

$$u_i = v_i, \quad \text{when} \quad i \neq 5, u_5 = v_5 + 2\sqrt{2}. \tag{2.129}$$

This transformation reduces the polynomial (2.128) to the canonical form

$$v_1^2 + v_2^2 + v_3^2 - 3v_4^2 + \sqrt{(2)}v_5. \tag{2.130}$$

Carrying out the transformations (2.123), (2.125), (2.127) and (2.129), we find a non-homogeneous orthogonal transformation which reduces the original polynomial to the canonical form (2.130),

$$x_1 = \frac{1}{2}(v_1 + v_2 + v_3 + v_4) - \frac{3}{2}, \quad x_2 = \frac{1}{2}(v_1 + v_2 - v_3 - v_4) - \frac{3}{2},$$

$$x_3 = \frac{1}{2}(v_1 - v_2 + v_3 - v_4) - \frac{1}{2}, \quad x_4 = \frac{1}{2}(v_1 - v_2 - v_3 + v_4) - \frac{1}{2},$$

$$x_5 = \frac{1}{2}\sqrt{2}(v_5 + v_6) + 2, \qquad x_6 = \frac{1}{2}\sqrt{2}(v_5 - v_6) + 2.$$

Orthogonal equivalence. Two real quadratic forms $A(\mathbf{x}; \mathbf{x})$ and $B(\mathbf{x}; \mathbf{x})$ in n variables $x_1, x_2, \ldots, x_n$ are said to be *orthogonally equivalent* if there exists an orthogonal transformation of the form (2.105) which, after it has been applied to the form $A(\mathbf{x}; \mathbf{x})$ and the variables y_i have again been replaced by the x_i, reduces the form $A(\mathbf{x}; \mathbf{x})$ to the form $B(\mathbf{x}; \mathbf{x})$. The orthogonal equivalence of forms is reflexive, symmetric and transitive (p. 120). Therefore, all real quadratic forms in n unknowns are separated into classes of orthogonally equivalent forms.

Two real quadratic forms are orthogonally equivalent if and only if their matrices have identical characteristic polynomials or, which is the same thing, identical characteristic roots ([20], Problem 1265). From this it follows that if in the diagonal form (2.109) we write the coefficients λ_i in a definite, for instance non-increasing, order, then every class of orthogonally equivalent quadratic forms (and this means also of orthogonally similar real symmetric matrices) is characterized uniquely by the form (respectively, matrix) of canonical form contained in it.

5. A pair of forms. A pencil of forms.

All the forms in this section are presupposed real. The properties of a pencil of quadratic forms has applications in the theory of small oscillations of mechanical systems (see [6], Chapter 10, section 8).

Suppose that we are given a pair of quadratic forms

$$A(\mathbf{x};\mathbf{x}) = \sum_{i,j=1}^{n} a_{ij}x_i x_j \quad \text{and} \quad B(\mathbf{x};\mathbf{x}) = \sum_{i,j=1}^{n} b_{ij}x_i x_j. \qquad (2.131)$$

The question arises, can both these forms be reduced to diagonal form by one non-singular linear transformation? In the general case this problem has no solution. Thus the forms

$$A(\mathbf{x},\mathbf{x}) = x_1^2, \quad B(\mathbf{x};\mathbf{x}) = x_1 x_2$$

cannot be reduced to diagonal form by one non-singular transformation (see [11], p. 231). We have the following sufficiency condition for the solution of the problem we have set.

If at least one of the two forms (2.131), for instance $B(\mathbf{x};\mathbf{x})$, is positive definite, then there exists a non-singular linear transformation of variables which reduces this form to normal form, and the other to diagonal form. To find such a transformation in practice it is more convenient first to find a non-singular transformation

$$\mathbf{X} = \mathbf{PY}, \qquad (2.132)$$

which reduces $B(\mathbf{x};\mathbf{x})$ to normal form

$$B_1(\mathbf{y};\mathbf{y}) = y_1^2 + y_2^2 + \ldots + y_n^2. \qquad (2.133)$$

Applying the same transformation (2.132) to the form $A(\mathbf{x};\mathbf{x})$, we obtain a new form

$$A_1(\mathbf{y};\mathbf{y}) = \sum_{i,j=1}^{n} a'_{ij}y_i y_j. \qquad (2.134)$$

For this form we find an orthogonal transformation of variables

$$\mathbf{Y} = \mathbf{QZ}. \qquad (2.135)$$

reducing it to diagonal form

$$A_2(\mathbf{z};\mathbf{z}) = \lambda_1 z_1^2 + \lambda_2 z_2^2 + \ldots + \lambda_n z_n^2. \qquad (2.136)$$

Applying the transformation (2.135) to the form (2.133), we obtain a form with matrix $\mathbf{Q'EQ'} = \mathbf{Q'Q} = \mathbf{E}$, i.e. once again a normal form

$$B_2(\mathbf{z};\mathbf{z}) = z_1^2 + z_2^2 + \ldots + z_n^2. \qquad (2.137)$$

Thus, the non-singular transformation of unknowns

$$\mathbf{X} = \mathbf{PQZ} \qquad (2.138)$$

reduces the form $A(\mathbf{x};\mathbf{x})$ to the diagonal form (2.136), and the form $B(\mathbf{x};\mathbf{x})$ to the normal form (2.137) ([11], p. 232).

A *pencil of quadratic forms* is a set of forms

$$A(\mathbf{x},\mathbf{x}) - \lambda B(\mathbf{x},\mathbf{x}), \tag{2.139}$$

where λ is a parameter taking any real values. If the form $B(\mathbf{x};\mathbf{x})$ is positive definite, then the pencil is called *regular*. The equation

$$|\mathbf{A} - \lambda\mathbf{B}| = 0 \tag{2.140}$$

is called the *characteristic equation of the pencil of forms* (2.139) (or its λ-equation), and its roots are called the characteristic roots or characteristic numbers of this pencil. If λ_0 is a characteristic root of the pencil (2.139), then there exists a non-zero vector $\mathbf{z} = (z_1, z_2, \ldots, z_n)$, the column of whose coordinates $\mathbf{Z}$ satisfies the equation

$$(\mathbf{A} - \lambda_0\mathbf{B})\mathbf{Z} = 0, \quad \mathbf{Z} \neq 0. \tag{2.141}$$

The vector $\mathbf{z}$ with this property is called the *principal vector of the pencil* (2.139), associated with the characteristic number λ_0.

The characteristic equation of a regular pencil (2.139) has n real characteristic roots (taking into account their multiplicity) $\lambda_1, \lambda_2, \ldots, \lambda_n$, with which are associated the principal vectors

$$\mathbf{z}_k = (z_{1k}, z_{2k}, \ldots, z_{nk}), \quad k = 1, 2, \ldots, n. \tag{2.142}$$

The columns $\mathbf{Z}_k$ of the coordinates of these vectors satisfy the conditions

$$\mathbf{A}\mathbf{Z}_k = \lambda_k\mathbf{B}\mathbf{Z}_k, \quad k = 1, 2, \ldots, n. \tag{2.143}$$

These principal vectors can be taken in such a way that the following conditions for ortho-normality are satisfied:

$$B(\mathbf{z}_i;\mathbf{z}_j) = \mathbf{Z}_i'\mathbf{B}\mathbf{Z}_j = \delta_{ij}, \quad i, j = 1, 2, \ldots, n. \tag{2.144}$$

Here $B(\mathbf{x},\mathbf{y})$ is the polar bilinear form of the quadratic form $B(\mathbf{x},\mathbf{x})$ ([6], p. 252, Theorem 8).

The *principal matrix* of the regular pencil (2.139) is the matrix

$$\mathbf{Z} = (z_{ij})_1^n, \tag{2.145}$$

whose columns consist of the coordinates of the principal vectors, satisfying the conditions (2.144). From this we can deduce that the columns of the principal matrix $\mathbf{Z}$ are linearly independent, and this means that this matrix is non-singular.

The non-singular linear transformation of unknowns

$$\mathbf{X} = \mathbf{ZY}, \tag{2.146}$$

where $\mathbf{X}$, $\mathbf{Y}$ are the columns of the old and the new variables and $\mathbf{Z}$ is the matrix of the transformation, reduces the positive definite form $\mathbf{B}(x, x)$ to a normal form, and the other form $A(x, x)$ to canonical form if and only if the matrix $\mathbf{Z}$ is the principal matrix of the regular pencil (2.139), corresponding to the given pair of forms. Here the coefficients λ_i of the canonical form of the form $A(x, x)$ are the characteristic roots (taking into account their multiplicity) of the pencil (2.139), hence, the canonical form is determined uniquely up to the order of the numbering of the coefficients λ_i ([6], p. 254, Theorem 9).

6. Extremal properties of the characteristic roots of a regular pencil of forms.

Suppose that we are given a regular pencil of quadratic forms (its definition and properties are given in the preceding section)

$$A(x, x) - \lambda B(x, x). \tag{2.147}$$

We arrange the characteristic roots of this pencil in non-decreasing order:

$$\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n. \tag{2.148}$$

We denote the corresponding principal vectors of the pencil by

$$z_k = (z_{1k}, z_{2k}, \ldots, z_{nk}); \quad k = 1, 2, \ldots, n. \tag{2.149}$$

We call two vectors x, y *orthogonal* if the equation $\mathbf{B}(x, y) = 0$ is satisfied for them. The principal vectors (2.149) can be regarded as pairwise orthogonal and non-zero (or even ortho-normal, i.e. satisfying equations (2.144) of the preceding section).

The smallest characteristic root λ_1 of a regular pencil of forms is equal to the minimum ratio of these forms:

$$\lambda_1 = \min \frac{A(x, x)}{B(x, x)}, \quad x \neq 0, \tag{2.150}$$

where this minimum is attained if and only if x is a principal vector of the pencil, associated with the value λ_1([6], p. 258, Theorem 10). The pth characteristic root λ_p, $2 \leqslant p \leqslant n$, is

equal to the minimum of the same ratio of the forms, where minimum is taken under the condition that the vector $\mathbf{x} \neq \mathbf{0}$ is orthogonal to the first $p-1$ principal vectors, orthogonal in pairs,

$$\left.\begin{aligned} &\lambda_p = \min \frac{A(\mathbf{x}, \mathbf{x})}{B(\mathbf{x}, \mathbf{x})}, \\ &\mathbf{x} \neq \mathbf{0}, \quad B(\mathbf{z}_i, \mathbf{x}) = 0, \quad i = 1, 2, \ldots, p-1, \\ &B(\mathbf{z}_i; \mathbf{z}_j) = 0, \quad i \neq j, \quad i, j = 1, 2, \ldots, p-1. \end{aligned}\right\} \tag{2.151}$$

This minimum is attained if and only if the vector x is a principal vector of the pencil, associated with the value λ_p([6], p. 259, Theorem 11).

For this characterization of the number λ_p it is necessary to know the first $p-1$ principal vectors, the choice of which in the general case is not unique. We can give another characterization for λ_p.

We say that h relations $L_1, L_2, \ldots, L_h$ have been imposed on the vector x (or on the variables $x_1, x_2, \ldots, x_n$), if its coordinates satisfy a system of h linear equations

$$L_k(\mathbf{x}) = l_{1k}x_1 + l_{2k}x_2 + \ldots + l_{nk}x_n = 0, \quad k = 1, 2, \ldots, h. \tag{2.152}$$

We will also consider special relations $\tilde{L}_k$, $k = 1, 2, \ldots, n$, given by the conditions of orthogonality of the vector x to the principal vectors:

$$\tilde{L}_k(\mathbf{x}) = B(\mathbf{z}_k, \mathbf{x}) = 0, \quad k = 1, 2, \ldots, n. \tag{2.153}$$

The minimum of the ratio of the forms

$$\frac{A(\mathbf{x}, \mathbf{x})}{B(\mathbf{x}, \mathbf{x})}$$

under the condition that the h relations $L_1, L_2, \ldots, L_h$ are imposed on the vector x, will be denoted by $\min\,(A/B;\ L_1, L_2, \ldots, L_h)$. Then for the pth characteristic root of the regular pencil of forms (2.147) we have the expression

$$\lambda_p = \max_{L_1, L_2, \ldots, L_{p-1}} \left[\min\left(\frac{A}{B};\ \tilde{L}_1, \tilde{L}_2, \ldots, \tilde{L}_{p-1}\right)\right], \tag{2.154}$$

where the minimum is taken with the given $p-1$ relations, and the maximum with a modification of these relations ([6], p. 260, Theorem 12).

Formula (2.151) can be rewritten thus:

$$\lambda_p = \min\left(\frac{A}{B};\ \tilde{L}_1, \tilde{L}_2, \ldots, \tilde{L}_{p-1}\right). \tag{2.155}$$

If we apply formulae (2.150), (2.155) and (2.154), after replacing $\mathbf{A}(x, x)$ by $-\mathbf{A}(x, x)$, we can obtain the formulae

$$\lambda_n = \max \frac{A(\mathbf{x}, \mathbf{x})}{B(\mathbf{x}, \mathbf{x})},\quad \mathbf{x} \neq \mathbf{0}; \tag{2.156}$$

$$\lambda_{n-p+1} = \max\left(\frac{A}{B};\ \tilde{L}_n, \tilde{L}_{n-1}, \ldots, \tilde{L}_{n-p+2}\right), \tag{2.157}$$

$$\lambda_{n-p+1} = \min_{L_1, L_2, \ldots, L_{p-1}}\left[\max\left(\frac{A}{B};\ L_1, L_2, \ldots, L_{p-1}\right)\right] \tag{2.158}$$

([6], p. 261, Theorem 13).

The relations $L_1, L_2, \ldots, L_h$ are called *independent* if the rank of the matrix of their coefficients is equal to h.

If we impose the h independent relations

$$L_1^0(\mathbf{x}) = 0,\ L_2^0(\mathbf{x}) = 0, \ldots,\quad L_h^0(\mathbf{x}) = 0, \tag{2.159}$$

on the regular pencil of forms (2.147) with characteristic roots (2.148), then from these equations h of the unknowns x_1, $x_2, \ldots, x_n$ can be expressed in terms of the remaining $n - h$. Substituting these expressions in the pencil (2.147), we obtain a new pencil

$$A^0(\mathbf{v}; \mathbf{v}) - \lambda B^0(\mathbf{v}; \mathbf{v}), \tag{2.160}$$

where we denote by $v_1, v_2, \ldots, v_{n-h}$ the remaining $n - h$ of the variables $x_1, x_2, \ldots, x_n$. The pencil (2.160) will again be regular. Its characteristic roots

$$\lambda_1^0 \leqslant \lambda_2^0 \leqslant \ldots \leqslant \lambda_{n-h}^0 \tag{2.161}$$

do not depend on which of the h previous unknowns x_i were excluded with the help of the given relations (2.159). Here the inequalities

$$\lambda_p \leqslant \lambda_p^0 \leqslant \lambda_{p+h};\ p = 1, 2, \ldots, n - h \tag{2.162}$$

are valid ([6], p. 262, Theorem 14).

Suppose that we are given two regular pencils of forms: the pencil (2.147) with characteristic roots (2.148) and the pencil

$$\tilde{A}(\mathbf{x}; \mathbf{x}) - \lambda \tilde{B}(\mathbf{x}; \mathbf{x}) \tag{2.163}$$

with characteristic roots

$$\tilde{\lambda}_1 \leqslant \tilde{\lambda}_2 \leqslant \ldots \leqslant \tilde{\lambda}_n. \tag{2.164}$$

If for any $x \neq 0$ the inequality

$$\frac{A(\mathbf{x}\,;\,\mathbf{x})}{B(\mathbf{x}\,;\,\mathbf{x})} \leqslant \frac{\tilde{A}(\mathbf{x}\,;\,\mathbf{x})}{\tilde{B}(\mathbf{x}\,;\,\mathbf{x})},$$

is satisfied, then $\lambda_p \leqslant \tilde{\lambda}_p$, $p = 1, 2, \ldots, n$ ([6], p. 263, Theorem 15).

If the second forms of the pencils (2.147) and (2.163) coincide and

$$\tilde{A}(\mathbf{x}\,;\,\mathbf{x}) = A(\mathbf{x}\,;\,\mathbf{x}) + \sum_{i=1}^{r} [X_i(\mathbf{x})]^2,$$

where $X_i(\mathbf{x})$, $i = 1, 2, \ldots, r$, are linearly independent linear forms, then

$$\lambda_p \leqslant \tilde{\lambda}_p \leqslant \lambda_{p+r}, \quad p = 1, 2, \ldots, n, \tag{2.165}$$

where the second parts of the inequalities only hold when $p \leqslant n - r$ ([6], p. 264, Theorem 16).

If the first forms of the pencils (2.147) and (2.163) coincide and

$$\tilde{B}(\mathbf{x}\,;\,\mathbf{x}) = B(\mathbf{x}\,;\,\mathbf{x}) + \sum_{i=1}^{r} [X_i(\mathbf{x})]^2,$$

where the $X_i(\mathbf{x})$, $i = 1, 2, \ldots, r$, are linearly independent linear forms, then the inequalities

$$\lambda_{p-r} \leqslant \tilde{\lambda}_p \leqslant \lambda_p, \quad p = 1, 2, \ldots, n, \tag{2.166}$$

hold, where the first parts of the inequalities only hold when $p > r$ ([6], p. 264, Theorem 17).

Taking as the second form in the pencil (2.147) the sum of squares

$$B(\mathbf{x}\,;\,\mathbf{x}) = x_1^2 + x_2^2 + \ldots + x_n^2,$$

from the properties of a regular pencil introduced in this section we obtain the extremal properties of the characteristic roots of one quadratic form $A(\mathbf{x}, \mathbf{x})$ (see [8], section 17).

7. Complex forms of the second kind. Hermitian forms.

In the case of the field of complex numbers, in addition to the bilinear and quadratic forms defined in subsection 1, we can consider bilinear and quadratic forms of another type.

Complex forms in the sense of subsection 1 will be called *forms of the first kind*, and in the new sense, *forms of the second kind*. This terminology is not customary, although it occurs in the applications to linear forms (see, for instance, [8], p. 79).

A *bilinear form of the second kind* is a function of $2n$ complex variables $x_1, x_2, \ldots, x_n$; $y_1, y_2, \ldots, y_n$, given by means of the expression

$$A(\mathbf{x}; \mathbf{y}) = \sum_{i,j=1}^{n} a_{ij} x_i \overline{y}_j, \tag{2.167}$$

where a_{ij} $(i, j = 1, 2, \ldots, n)$ are any complex numbers, and the bar over y_j means that in substituting the values of the arguments, instead of $\overline{y}_j$ we substitute the number which is the complex conjugate of y_j. The coefficients a_{ij}, and hence also the expression for the bilinear form in the form of the sum (2.167), are defined uniquely for a given form.

The matrix of the coefficients a_{ij}

$$\mathbf{A} = (a_{ij})_1^n \tag{2.168}$$

is called the *matrix of the bilinear form* (2.167), and its rank the *rank of this form*. We can define a bilinear form of the second kind as a function of two vectors $\mathbf{x}$, $\mathbf{y}$ of complex linear space (see [8], p. 80). We will assume that the vectors $\mathbf{x}$, $\mathbf{y}$ are given by their coordinates $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)$.

A *quadratic form of the second kind* is a function of n complex variables $x_1, x_2, \ldots, x_n$, obtained from a bilinear form of the second kind when $x = y$ or when $x_i = y_i$, $i = 1, 2, \ldots, n$, i.e. a function given by means of the expression

$$A(\mathbf{x}; \mathbf{x}) = \sum_{i,j=1}^{n} a_{ij} x_i \overline{x}_j. \tag{2.169}$$

As distinct from quadratic forms of the first kind, the coefficients a_{ij} and thus also the representation (2.169), are uniquely determined by the quadratic form of the second kind.

The matrix, rank, discriminant, the concept of a non-degenerate and a degenerate (or singular) quadratic form for a form of the second kind are defined by analogy with the case of forms of the first kind (pp. 118–119).

Because bilinear and quadratic forms of the second kind lie in one–one correspondence with their matrices, then as

distinct from forms of the first kind, we can establish a one–one correspondence between bilinear and quadratic forms of the second kind, where the forms corresponding to each other have identical matrices. Therefore, the values of a bilinear form of the second kind must be defined by the values of the corresponding quadratic form. There is the following connexion between these forms:

$$A(\mathbf{x};\mathbf{y}) = \frac{1}{4}[A(\mathbf{x}+\mathbf{y};\mathbf{x}+\mathbf{y}) + iA(\mathbf{x}+i\mathbf{y};\mathbf{x}+i\mathbf{y}) - A(\mathbf{x}-\mathbf{y};\mathbf{x}-\mathbf{y}) - iA(\mathbf{x}-i\mathbf{y};\mathbf{x}-i\mathbf{y})], \quad (2.170)$$

$$A(\mathbf{y};\mathbf{x}) = \frac{1}{4}[A(\mathbf{x}+\mathbf{y};\mathbf{x}+\mathbf{y}) - iA(\mathbf{x}+i\mathbf{y};\mathbf{x}+i\mathbf{y}) - A(\mathbf{x}-\mathbf{y};\mathbf{x}-\mathbf{y}) + iA(\mathbf{x}-i\mathbf{y};\mathbf{x}-i\mathbf{y})] \quad (2.171)$$

([8], pp. 81–82, (1), (2)).

By means of the multiplication of matrices a bilinear form of the second kind (2.167) can be rewritten in the form

$$A(\mathbf{x};\mathbf{y}) = \mathbf{X}'\mathbf{A}\overline{\mathbf{Y}}, \quad (2.172)$$

and the quadratic form of the second kind (2.169) in the form

$$A(\mathbf{x};\mathbf{x}) = \mathbf{X}'\mathbf{A}\overline{\mathbf{X}}, \quad (2.173)$$

where $\mathbf{X}$, $\mathbf{Y}$ are the columns of the variables x_i, y_i. If we apply the transformation of variables

$$\mathbf{X} = \mathbf{SU}, \quad \mathbf{Y} = \mathbf{TV}, \quad (2.174)$$

to the form (2.172), then its matrix is transformed by the formula

$$\mathbf{B} = \mathbf{S}'\mathbf{A}\overline{\mathbf{T}}. \quad (2.175)$$

The matrix of the quadratic form (2.173) under the transformation of variables

$$\mathbf{X} = \mathbf{TY} \quad (2.176)$$

is changed by means of the formula

$$\mathbf{B} = \mathbf{T}'\mathbf{A}\overline{\mathbf{T}}. \quad (2.177)$$

After we make the substitution $\mathbf{Q} = \overline{\mathbf{T}}$, this formula takes the form

$$\mathbf{B} = \mathbf{Q}^*\mathbf{A}\mathbf{Q}, \quad \mathbf{Q}^* = \overline{\mathbf{Q}}'. \quad (2.178)$$

From the formulae (2.175) and (2.177) it follows that under a non-singular linear transformation of variables the rank of a bilinear and a quadratic form of the second kind is unaltered.

Amongst forms of the second kind, a part similar to the symmetric real bilinear and real quadratic forms is played by the so-called Hermitian forms. All the fundamental results of subsections 1–5 carry over to these forms (where everywhere the scalar multiplication of vectors must be defined by the formula (2.52), p. 103) (see [6], Chapter 10, section 9).

A *Hermitian* bilinear (or quadratic) form is a bilinear (respectively, quadratic) form of the second kind whose matrix is Hermitian (p. 109), i.e. possesses the property

$$\mathbf{A}^* = \mathbf{A}, \quad a_{ij} = \bar{a}_{ji}, \quad i, j = 1, 2, \ldots, n. \tag{2.179}$$

From equation (2.178) it follows that Hermitian bilinear and quadratic forms under a linear transformation of variables (2.176) are carried again into Hermitian forms.

A bilinear form of the second kind $A(\mathbf{x}; \mathbf{y})$ will be Hermitian if and only if the identity

$$A(\mathbf{x}; \mathbf{y}) = \overline{A(\mathbf{y}; \mathbf{x})} \tag{2.180}$$

is satisfied. From this and from equations (2.170) and (2.171) it follows that a quadratic form of the second kind $A(\mathbf{x}; \mathbf{x})$ will be Hermitian if and only if it has real values for any complex values of $x_1, x_2, \ldots, x_n$.

From the conditions (2.179) for the matrix $\mathbf{A}$ to be Hermitian it follows that its elements lying along the principal diagonal and in general all the principal minors $A\begin{pmatrix} i_1 i_2 \ldots i_k \\ i_1 i_2 \ldots i_k \end{pmatrix}$ are real.

The methods of Lagrange and Jacobi for the reduction to diagonal form (subsection 1) carry over to Hermitian quadratic forms. In excluding the first term by Lagrange's method when $a_{11} \neq 0$ we gather together all the summands containing x_1 and $\bar{x}_1$, and apply the identity

$$a_{11} x_1 \bar{x}_1 + \sum_{k=2}^{n} a_{1k} x_1 \bar{x}_k + \sum_{k=2}^{n} a_{k1} x_k \bar{x}_1 = \frac{1}{a_{11}} \left(\sum_{k=1}^{n} a_{1k} x_k \right) \cdot \left(\sum_{k=1}^{n} \bar{a}_{1k} \bar{x}_k \right)$$

$$- \frac{1}{a_{11}} \left(\sum_{k=2}^{n} a_{1k} x_k \right) \cdot \left(\sum_{k=2}^{n} a_{1k} \bar{x}_k \right).$$

Thus, any Hermitian quadratic form can be reduced to the diagonal form

$$\lambda_1 y_1 \bar{y}_1 + \lambda_2 y_2 \bar{y}_2 + \ldots + \lambda_n y_n \bar{y}_n, \tag{2.181}$$

by means of a non-singular complex linear transformation of variables, where the coefficients λ_i are real; for forms of rank r we can assume that $\lambda_i \neq 0$, when $i = 1, 2, \ldots, r$; $\lambda_i = 0$, when $i = r + 1, \ldots, n$.

For Hermitian quadratic forms there remain in force the law of inertia, the concepts of the indices of inertia and of the signature, their definition by means of the signs of the angular minors, Frobenius's and Hildenfinger's rules (subsection 2), and also the definition and criteria for positive definite forms (subsection 3).

From the fact that any Hermitian matrix is unitarily similar to a real diagonal matrix with characteristic numbers along the diagonal (p. 111) there follows

Reduction of a Hermitian quadratic form to principal axes. Any Hermitian quadratic form (2.169) with matrix **A** can be reduced to diagonal form (2.181) by means of a unitary transformation of variables of the form (2.176). The coefficients λ_i of this form will be the characteristic numbers of the matrix **A** and, hence, real and uniquely determined. The columns of the matrix of the transformation **T** consist of numbers which are the complex conjugates of the coordinates of an ortho-normal system of characteristic vectors of the matrix **A**. We note that the ortho-normal system of vectors must be understood in the sense of scalar multiplication by formula (2.52) (p. 103).

The properties of pairs of forms (subsection 5) and the extremal properties of the characteristic numbers of a regular pencil (subsection 6) also carry over to any Hermitian forms (see [6], Chapter 10, section 9).

8. Hänkel forms.

Suppose that we are given a sequence of $2n - 1$ numbers $s_0, s_1, \ldots, s_{2n-2}$. A *Hänkel form* is a quadratic form in n variables $x_0, x_1, \ldots, x_{n-1}$ of the form

$$S(\mathbf{x}; \mathbf{x}) = \sum_{i,j=0}^{n-1} s_{i+j} x_i x_j. \tag{2.182}$$

The symmetric matrix of the Hänkel form

$$S = \begin{pmatrix} s_0 & s_1 & s_2 & \dots & s_{n-1} \\ s_1 & s_2 & s_3 & \dots & s_n \\ s_2 & s_3 & s_4 & \dots & s_{n+1} \\ \dots & \dots & \dots & \dots & \dots \\ s_{n-1} & s_n & s_{n+1} & \dots & s_{2n-2} \end{pmatrix} \tag{2.183}$$

is called a *Hänkel matrix.*

For real Hänkel forms the indices of inertia and the signature are completely determined by the values of the principal minors of their matrices. The following three propositions are true for Hänkel forms with coefficients from any field.

(1) If in the Hänkel matrix (2.183) the first h rows are linearly independent, and the first $h-1$ are linearly dependent $(h < n)$, then

$$D_h = S\begin{pmatrix} 1 & 2 & \dots & h \\ 1 & 2 & \dots & h \end{pmatrix} \neq 0 \tag{2.184}$$

([6], p. 274, Lemma 1).

(2) If for the Hänkel matrix (2.183) the conditions

$$D_h \neq 0, \quad D_{h+1} = \dots = D_n = 0, \quad h < n, \tag{2.185}$$

are satisfied for some h, and if we set

$$t_{ij} = \frac{1}{D_h} S\begin{pmatrix} 1 & 2 & \dots & h & h+i+1 \\ 1 & 2 & \dots & h & h+j+1 \end{pmatrix}, \quad i, j = 0, 1, \dots, n-h-1, \tag{2.186}$$

then the matrix $\mathbf{T} = (t_{ij})_0^{n-h-1}$ is also Hänkel, and all its elements lying above its secondary diagonal are equal to zero ([6], p. 274, Lemma 2).

(3) If the Hänkel matrix (2.183) has rank r and for some h the conditions

$$D_h \neq 0, \quad D_{h+1} = \dots = D_r = 0, \quad h < r, \tag{2.187}$$

are satisfied, then the principal minor $D^{(r)}$ of order r, which stands at the intersections of the first h and the last $r-h$ rows and the same columns, is distinct from zero, i.e.

$$D^{(r)} = S\begin{pmatrix} 1 & 2 & \dots & h & n-r+h+1 & \dots & n \\ 1 & 2 & \dots & h & n-r+h+1 & \dots & n \end{pmatrix} \neq 0 \tag{2.188}$$

([6], p. 276, Theorem 23).

(4) *Frobenius's theorem.* For a real Hänkel form (2.182) of rank r the indices of inertia p, q and the signature s can be defined by the signs of the angular minors (cf. subsection 2) by means of the formulae

$$\left.\begin{aligned} p &= P(1, D_1, D_2, \ldots, D_r), \\ q &= V(1, D_1, D_2, \ldots, D_r), \\ s &= r - 2V(1, D_1, D_2, \ldots, D_r), \end{aligned}\right\} \tag{2.189}$$

where P and V denote, respectively, the number of unchanged signs and the number of changes of sign in the series of numbers in the brackets.

Here it is presupposed that $\mathbf{D}_r \neq \mathbf{0}$ and that in any group of m intermediate minors

$$D_h \neq 0, \quad D_{h+1} = \ldots = D_{h+m} = 0, \quad D_{h+m+1} \neq 0 \tag{2.190}$$

the zero minors can be ascribed signs by means of the formula

$$\operatorname{sign} D_{h+j} = (-1)^{\frac{j(j-1)}{2}} \operatorname{sign} D_h. \tag{2.191}$$

With such a choice of signs, the quantities P, V, $P - V$, corresponding to the group of minors (2.190), take the values:

$$P = \begin{cases} \dfrac{m+1}{2} & \text{when } m \text{ is odd,} \\ \dfrac{m+1+\varepsilon}{2} & \text{when } m \text{ is even,} \end{cases}$$

$$V = \begin{cases} \dfrac{m+1}{2} & \text{when } m \text{ is odd,} \\ \dfrac{m+1-\varepsilon}{2} & \text{when } m \text{ is even,} \end{cases}$$

$$P - V = \begin{cases} 0 & \text{when } m \text{ is odd,} \\ \varepsilon & \text{when } m \text{ is even} \end{cases}$$

$$\varepsilon = (-1)^{\frac{m}{2}} \operatorname{sign} \frac{\mathbf{D}_{n+m+1}}{\mathbf{D}_n}.$$

In the case when $\mathbf{D}_r = \mathbf{0}$, what we have said remains true, if we replace $\mathbf{D}_r$ by $\mathbf{D}^{(r)} \neq \mathbf{0}$ according to (2.188) ([6], p. 277, Theorem 24).

CHAPTER III

THE ALGEBRA OF POLYNOMIALS[11]

1. The General Properties of Polynomials

1. Definitions. Examples.

A *polynomial of degree n* (n is a non-negative integer) in the variable x over the number field K (see Introduction) is an expression of the form

$$a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n,$$

where $a_0, \ldots, a_n$ are numbers from K, and $a_0 \neq 0$.† A polynomial of zero degree is simply a number distinct from zero. The number zero is also regarded as a polynomial; its degre is not defined.

The usual notation for polynomials is $f(x)$, $g(x)$, $\varphi(x)$ and so on. Two polynomials are considered *equal* if their coefficients for the same powers of the variable are equal.

Polynomials can be added, subtracted, and multiplied by each other. The *sum* (*difference*) of two polynomials $f(x)$ and $g(x)$ is the polynomial whose coefficient for each power of the variable x is equal to the sum (difference) of the coefficients of the same power of x in the polynomials $f(x)$ and $g(x)$.

EXAMPLE 1.

$$f(x) = x^3 - 4x^2 + x + 1, \quad g(x) = x^2 + 1;$$

$$f(x) + g(x) = x^3 - 3x^2 + x + 2; \quad f(x) - g(x) = x^3 - 5x^2 + x.$$

However, to *multiply* the polynomial $f(x)$ by the polynomial $g(x)$, it is necessary to multiply every term of the polynomial $f(x)$ by every term of $g(x)$, add the products obtained and collect similar terms. The degree of the product of two polynomials is always equal to the sum of the degrees of the factors.

† In a similar way we define a polynomial over an arbitrary field P, not necessarily with numerical elements (Chapter IV, section 3).

EXAMPLE 2.

$$f(x) = x^2 - 3x + 2, \quad g(x) = x^2 + 2x - 3;$$

$$f(x)g(x) = x^4 + 2x^3 - 3x^2 - 6x^2 + 9x + 2x^2 + 4x - 6$$

$$= x^4 - x^3 - 7x^2 + 13x - 6.$$

The (exact) division of one polynomial by another is not always possible.

2. Division with a remainder.

Any polynomial $f(x)$ can be *divided* by any other polynomial $g(x)$ $(g(x) \neq 0)$ *with a remainder*, i.e. it is possible to represent the polynomial $f(x)$ in a unique way in the form $f(x) = g(x)q(x) + r(x)$, where $q(x)$ and $r(x)$ are polynomials, and the degree of $r(x)$ is less than the degree of $g(x)$ or else $r(x) = 0$; $q(x)$ is called the *quotient* in the division of the polynomial $f(x)$ by $g(x)$, $r(x)$ is called the *remainder* in this division. If $r(x) = 0$, then we say that $f(x)$ is (exactly) *divisible by* $g(x)$ *without remainder*.

If $g(x)$ is a polynomial of zero degree, i.e. a number, then any polynomial $f(x)$ is divisible by it without remainder.

$f(x)$ is divisible without remainder by $g(x)$ and at the same time $g(x)$ is divisible without remainder by $f(x)$ if and only if $g(x) = cf(x)$, where c is some number different from zero.

EXAMPLE 1.

$$f(x) = x^3 - 1, \quad g(x) = x^2 + x + 1.$$

Here $x^3 - 1 = (x^2 + x + 1)(x - 1)$, i.e. $f(x)$ is divisible by $g(x)$ without remainder $(q(x) = x - 1, \ r(x) = 0)$.

EXAMPLE 2.

$$f(x) = 2x^2 - 3x + 1, \quad g(x) = x^2 - 4x - 1.$$

Here

$$f(x) = g(x)\,q(x) + r(x), \quad \text{where} \quad q(x) = 2, \quad r(x) = 5x + 3.$$

If the degree of $f(x)$ is less than the degree of $g(x)$, then the quotient after the division of $f(x)$ by $g(x)$ is the polynomial $q(x) = 0$, and the remainder is the polynomial $f(x)$ itself. In the general case, if finding the quotient and remainder we can make use of the *rule for division "by the corner"* learnt at school (it applies to polynomials with coefficients from any field).

3. Horner's scheme.

If the polynomial $f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$ is to be divided by the polynomial $g(x) = x - b$, then it is easiest to do means of *Horner's scheme*, which is formed thus: write

down in a row the coefficients of the polynomial $f(x)$, beginning with the number a_0; on the left write the number b: then below a_0 write the number a_0 again, below a_1 write the number $a_0b + a_1 = b_1$, below a_2 the number $b_1b + a_2 = b_2$ and so on; below a_n write the number $b_{n-1}b + a_n = b_n$.

b_n is the remainder after the division of $f(x)$ by $x - b$, $a_0, b_1, \ldots, b_{n-1}$ are the coefficients of the quotient.

EXAMPLE 1.

$$f(x) = 2x^3 - 3x + 5, \quad g(x) = x - 4.$$

We form Horner's scheme:

	2	0	− 3	5
4	2	8	29	121

$$f(x) = (x - 4)(2x^2 + 8x + 29) + 121.$$

If we wish to divide the polynomial $f(x)$ by the polynomial $g(x) = ax + b$, $a \neq 0$, then we can begin by dividing it by Horner's method by the polynomial $g_1(x) = x - (-b/a)$; if $q(x)$ and r are the quotient and the remainder after the division of $f(x)$ by $g(x)$, and $q_1(x)$, r_1 the quotient and remainder after the division of $f(x)$ by $g_1(x)$, then $r = r_1$, and $q(x) = (1/a)\, q_1(x)$.

EXAMPLE 2.

$$f(x) = x^3 - 2x^2 - 10x + 3, \; g(x) = 2x + 5.$$

We take $g_1(x) = x + \frac{5}{2}$ and we write out Horner's scheme for the division of $f(x)$ by $g_1(x)$:

	1	− 2	− 10	3
$-\frac{5}{2}$	1	$-\frac{9}{2}$	$\frac{5}{4}$	$-\frac{1}{8}$

We obtain $f(x) = g(x)q(x) + r$, where

$$q(x) = \frac{1}{2}\left(x^2 - \frac{9}{2}x + \frac{5}{4}\right) = \frac{1}{2}x^2 - \frac{9}{4}x + \frac{5}{8}, \quad r = -\frac{1}{8}.$$

The remainder after the division of an arbitrary polynomial $f(x)$ by the polynomial $g(x) = x - b$ is a number, equal to the value of the polynomial $f(x)$ when $x = b$ (Bezout's theorem). Because of this, Horner's method is also convenient for the quick calculation of the value of a polynomial for a given value of the unknown.

EXAMPLE 3. $f(x) = x^4 - 2x^3 + x^2 + x + 1$. To find the value of $f(x)$ when $x = -3$.

We form Horner's scheme:

	1	− 2	1	1	1
− 3	1	− 5	17	− 47	142

$$f(-3) = 142.$$

4. The expansion of a polynomial into powers of a difference ([23], pp. 97–8; [31], pp. 77–9).

For any polynomial

$$f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$$

and any number c we can write out the *expansion of* $f(x)$ *into powers of the difference* $x - c$:

$$f(x) = b_0(x-c)^n + b_1(x-c)^{n-1} + \ldots + b_{n-1}(x-c) + b_n.$$

In order to find the coefficients of this expansion, it is necessary first to divide $f(x)$ by $x - c$. As the remainder we obtain b_n, and the quotient will be some polynomial $q(x)$. Then we must divide $q(x)$ by $x - c$; as the remainder we obtain b_{n-1}, as the quotient, $q_1(x)$. Then we divide $q_1(x)$ by $x - c$, as the remainder we obtain b_{n-2} and so on. It is convenient to carry out the calculation by Horner's scheme (subsection 3).

EXAMPLE 1.

$$f(x) = x^4 - 5x^3 - 3x^2 + 9, \quad c = 3.$$

We form Horner's scheme; we write out in the first row the coefficients of the polynomial $f(x)$, in the second we obtain the coefficients of the quotient $q(x)$ and the remainder b_4 after the division of $f(x)$ by $x - 3$, in the third row we obtain the coefficients of the quotient $q_1(x)$ and the remainder b_3 after the division of $q(x)$ by $x - 3$ and so on:

	1	−5	−3	0	9
3	1	−2	−9	−27	−72
	1	1	−6	−45	
	1	4	6		
	1	7			
	1				

$$f(x) = (x-3)^4 + 7(x-3)^3 + 6(x-3)^2 - 45(x-3) - 72.$$

The coefficients of the expansion of the polynomial $f(x)$ into powers of the difference $x - c$ are connected with the values of this polynomial and its derivatives† when $x = c$:

$$b_n = f(c), \quad b_{n-1} = \frac{f'(c)}{1!}, \quad b_{n-2} = \frac{f''(c)}{2!}, \ldots, b_0 = \frac{f^{(n)}(c)}{n!}.$$

† The derivative of the polynomial $f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$ with coefficients from any field is the polynomial

$$f'(x) = na_0x^{n-1} + (n-1)a_1x^{n-2} + \ldots + a_{n-1}.$$

Therefore, Horner's scheme, written out in the preceding example, also enables us to find the values of the derivatives of the polynomial $f(x)$ when $x = c$ (in the example cited these values are: $f'(3) = -45$, $f''(3) = 6 \cdot 2! = 12$, $f'''(3) = 7 \cdot 3! = 42$, $f^{IV}(3) = 4! = 24$).

The expansion of a polynomial into powers of the difference $x - c$ can be used in the expansion of a fraction whose denominator is the power of a linear binomial, into partial fractions (section 3, subsection 9).

EXAMPLE 2.

$$\Phi(x) = \frac{x^2 + x + 1}{(x+2)^4}.$$

We take the polynomial $f(x) = x^2 + x + 1$ and expand it in powers of the difference $x - (-2) = x + 2$:

$$\begin{array}{r|rrr} & 1 & 1 & 1 \\ \hline -2 & 1 & -1 & 3 \\ & 1 & -3 & \\ \hline & 1 & & \end{array}$$

$$f(x) = (x+2)^2 - 3(x+2) + 3.$$

We obtain

$$\Phi(x) = \frac{f(x)}{(x+2)^4} = \frac{1}{(x+2)^2} - \frac{3}{(x+2)^3} + \frac{3}{(x+2)^4}.$$

5. The greatest common divisor of two polynomials.

If each of the polynomials $f(x)$ and $g(x)$ are divisible (without remainder) by the polynomial $\varphi(x)$, then $\varphi(x)$ is called a *common divisor* of $f(x)$ and $g(x)$. A *greatest common divisor* (g.c.d.) of the polynomials $f(x)$ and $g(x)$ is a common divisor $d(x)$ of them which is divisible by all other common divisors of the polynomial $f(x)$ and $g(x)$ (notation: $d(x) = (f(x), g(x))$. For any two polynomials not identically equal to zero, a g.c.d. exists and is defined uniquely to within constant factors distinct from zero. Of all the greatest common divisors of the polynomials $f(x)$ and $g(x)$ we usually choose the one whose leading coefficient is 1.

EXAMPLE 1.

$$f(x) = x^2 - ix, \quad g(x) = x^2 + 1, \quad d(x) = x - i.$$

EXAMPLE 2.

$$f(x) = 6x + 12, \quad g(x) = 4x - 2, \quad d(x) = 1.$$

If we know the decompositions of the polynomials $f(x)$ and $g(x)$ into linear factors (section 3, subsection 6), where

$$f(x) = a_0(x-\alpha_1)^{k_1}\dots(x-\alpha_p)^{k_p}(x-\beta_1)^{u_1}\dots(x-\beta_t)^{u_t};$$

$$g(x) = b_0(x-\alpha_1)^{l_1}\dots(x-\alpha_p)^{l_p}(x-\gamma_1)^{v_1}\dots(x-\gamma_s)^{v_s}$$

(the numbers $\alpha_j, \beta_q, \gamma_v$ are all distinct from each other), then $d(x) = (x-\alpha_1)^{m_1}\dots(x-\alpha_p)^{m_p}$, where for every j ($j = 1, \dots, \rho$) m_j is the smaller of the numbers k_j, l_j.

EXAMPLE 3.

$$f(x) = (x-1)^3(x+2)^2(x-5)^3,$$

$$g(x) = (x-1)(x+2)^4(x+7)(x+1)^2; \quad d(x) = (x-1)(x+2)^2.$$

If the decompositions of the polynomials $f(x)$ and $g(x)$ into linear factors are not known, then to find their g.c.d. we apply *Euclid's algorithm*. The polynomial $f(x)$ is divided by $g(x)$ (subsection 2); we obtain the remainder $r_1(x)$ (possibly equal to zero); then, if $r_1(x) \neq 0$, $g(x)$ is divided by $r_1(x)$; we obtain the remainder $r_2(x)$; if $r_2(x) \neq 0$, then $r_1(x)$ is divided by $r_2(x)$, and so on, until we obtain a remainder equal to zero. The last remainder not equal to zero, $r_k(x)$, is the g.c.d. of the polynomials $f(x)$ and $g(x)$. If, however, $r_1(x) = 0$, then the g.c.d. of the polynomials $f(x)$ and $g(x)$ is $g(x)$.

EXAMPLE 4.

$$f(x) = 3x^3 - 2x^2 + x + 4, \quad g(x) = x^2 - x - 1.$$

We divide $f(x)$ by $g(x)$:

$$f(x) = g(x)(3x+1) + 5(x+1).$$

We obtain $r_1(x) = 5(x+1) \neq 0$. We divide $g(x)$ by $r_1(x)$:

$$g(x) = r_1(x)\frac{1}{5}(x-2) + 1; \quad r_2(x) = 1 \neq 0.$$

$r_1(x)$ is divisible by $r_2(x)$ without remainder. Thus, $(f(x), g(x)) = 1$.

By applying Euclid's algorithm we can multiply the dividend and divisor by any non-zero numbers, and this can even be done in the process of the division itself. In this way we are able to perform operations on polynomials with integral coefficients without having to deal with fractions.

EXAMPLE 5.

$$f(x) = 2x^4 + 2x^3 + x^2 - x - 1, \quad g(x) = 3x^4 + 2x^2 - x + 2.$$

We divide the polynomial $3f(x)$ (in the usual way) by $g(x)$ and we obtain as the remainder the polynomial $r_1(x) = 6x^3 - x^2 - x - 7$. Then we divide $2g(x)$ by $r_1(x)$:

$$\begin{array}{rl|l}
6x^4 + 4x^3 - 2x + 4 & & 6x^3 - x^2 - x - 7 \\
\cline{3-3}
6x^4 - x^3 - x^2 - 7x & & x + 1 \\
\cline{1-1}
x^3 + 5x^2 + 5x + 4 & & \\
\cline{1-1}
6x^3 + 30x^2 + 30x + 24 & & \\
6x^3 - x^2 - x - 7 & & \\
\cline{1-1}
31x^2 + 31x + 31 & &
\end{array}$$

(Here in the process of division we have performed a multiplication by 6; the symbol || separates the part of the quotient in which the alteration occurred.) The remainder (after division by 31) is equal to

$$r_2(x) = x^2 + x + 1;$$

$r_1(x)$ is divisible by $r_2(x)$ without remainder, hence, $r_2(x) = (f(x), g(x))$.

6. Mutually prime polynomials.

Two polynomials whose g.c.d. is equal to 1 are said to be *mutually prime*. Polynomials are mutually prime if and only if they do not have a common root (subsection 10).

EXAMPLE.

$$f(x) = x^3 - 3x^2 + 5, \quad g(x) = x^2 - 4x + 3.$$

Here $d(x) = 1$, the polynomials $f(x)$ and $g(x)$ are mutually prime.

If $d(x) = (f(x), g(x))$, then the polynomials $f_1(x) = f(x)/d(x)$ and $g_1(x) = g(x)/d(x)$ are mutually prime. If the polynomial $\varphi(x)$ is mutually prime to each of the polynomials $f(x)$ and $g(x)$, then it is mutually prime to their product. If the product of the polynomials $f(x)$ and $g(x)$ is divisible by $\varphi(x)$, and $f(x)$ and $\varphi(x)$ are mutually prime, then $g(x)$ is divisible by $\varphi(x)$. If the polynomial $f(x)$ is divisible by each of the polynomials $\varphi(x)$ and $\psi(x)$, which are mutually prime to each other, then $f(x)$ is also divisible by their product.

7. The representation of $d(x) = (f(x), g(x))$ in the form $d(x) = f(x)\,u(x) + g(x)\,v(x)$.

The g.c.d. $d(x)$ of the polynomials $f(x)$ and $g(x)$ can always be represented in the form

$$d(x) = f(x)u(x) + g(x)v(x), \tag{3.1}$$

where $u(x)$ and $v(x)$ are polynomials, and this can be done in an uncountable number of different ways. If, in particular, the polynomials $f(x)$ and $g(x)$ are mutually prime (subsection 6), then we obtain $1 = f(x)u(x) + g(x)v(x)$.

If the degrees of the polynomials $f(x)$ and $g(x)$ are greater than zero, then the polynomials $u(x)$ and $v(x)$ in equation (3.1) can be selected in such a way that the degree of $u(x)$ will be less than the degree of $g(x)$, and the degree of $v(x)$ less than the degree of $f(x)$. In practice, when finding the polynomials $u(x)$, and $v(x)$ it is more convenient to take, instead of $f(x)$ and $g(x)$, the polynomials $f_1(x) = f(x)/d(x)$, $g_1(x) = g(x)/d(x)$ and to select $u(x)$, $v(x)$ first in such a way that the equation

$$1 = f_1(x)u(x) + g_1(x)v(x) \tag{3.2}$$

is satisfied. This is done in the following way. If $f_1(x)$ or $g_1(x)$ is a polynomial of zero degree, then the polynomials $u(x)$ and $v(x)$ are easy to choose: for instance, when $g_1(x) = a \neq 0$ we can set $u(x) = 0$, $v(x) = 1/a$. If, however, the degrees of $f_1(x)$ and $g_1(x)$ are positive, then it is necessary to write $u(x)$ and $v(x)$ with indeterminate coefficients (assuming that the degree of $u(x)$ is less than the degree of $g_1(x)$, and the degree of $v(x)$ less than the degree of $f_1(x)$) and to equate the coefficients of similar powers of x on the left- and right-hand sides of equation (3.2) (or else we can give the variable x different numerical values and equate to 1 the values of the polynomial $f_1(x)u(x) + g_1(x)v(x)$ obtained in this way); this gives a system of linear equations whose solution (Chapter I, section 2, subsections 1–3) enables us to find the unknown coefficients of the polynomials $u(x)$ and $v(x)$. If we now multiply both sides of equation (3.2) by $d(x)$, then we obtain equation (3.1) (with the same polynomials $u(x)$ and $v(x)$ as in (3.2)). In this method of finding $u(x)$ and $v(x)$ the degree of $u(x)$ obtained is smaller than the difference between the degrees of $g(x)$ and $d(x)$, and the degree of $v(x)$ smaller than the difference between the degrees of $f(x)$ and $d(x)$ (provided that the given difference is non-zero). With these conditions on the degrees, the polynomials $u(x)$ and $v(x)$ in equation (3.1) are defined uniquely.

EXAMPLE 1.

$$f(x) = x^7 - 2x^6 - x^5 - 11x^4 - 6x^3 - 4x^2 + 5x + 14, \quad g(x) = x^2 + x + 2.$$

Here $d(x) = g(x) = x^2 + x + 2$, hence, we can take $u(x) = 0$, $v(x) = 1$.

EXAMPLE 2.

$$f(x) = 3x^5 - 4x^4 + x^3 - 3x^2 + 4x - 1, \quad g(x) = 3x^5 + 5x^4 + x^3 - x^2 - 3x + 1.$$

Here $d(x) = 3x^3 + 2x^2 + 2x - 1$, $f_1(x) = x^2 - 2x + 1$, $g_1(x) = x^2 + x - 1$. For the polynomials $f_1(x)$, $g_1(x)$ we write equation (3.2) with indeterminate coefficients for $u(x)$ and $v(x)$:

$$1 = (x^2 - 2x + 1)(ax + b) + (x^2 + x - 1)(a_1x + b_1).$$

We equate coefficients of similar powers of x and obtain the system of equations

$$\begin{cases} a + a_1 = 0, \\ b - 2a + b_1 + a_1 = 0, \\ -2b + a + b_1 - a_1 = 0, \\ b - b_1 = 1. \end{cases}$$

Solving this system, we find: $a = 3$, $a_1 = -3$, $b = 5$, $b_1 = 4$, i.e.

$$d(x) = f(x)(3x + 5) + g(x)(-3x + 4).$$

EXAMPLE 3.

$$f(x) = 2x^4 - 4x^3 - 3x^2 + 7x - 2, \quad g(x) = 6x^3 + 4x^2 - 5x + 1.$$

Here $d(x) = 2x^2 + 2x - 1$, $f_1(x) = x^2 - 3x + 2$, $g_1(x) = 3x - 1$. We write equation (3.2):

$$1 = (x^2 - 3x + 2)a + (3x - 1)(bx + c).$$

Giving x the values 0, 1, 2, in succession, we obtain the system of equations

$$\begin{cases} 2a - c = 1, \\ 2b + 2c = 1, \\ 10b + 5c = 1. \end{cases}$$

Solving this system, we find: $a = 0{\cdot}9$, $b = -0{\cdot}3$, $c = 0{\cdot}8$, whence $d(x) = f(x)0{\cdot}9 + g(x)(-0{\cdot}3x + 0{\cdot}8)$.

In order to represent a polynomial $h(x)$. distinct from the g.c.d. $d(x)$ of the polynomials $f(x)$ and $g(x)$, in the form

$$h(x) = f(x)u_1(x) + g(x)v_1(x), \tag{3.3}$$

where $u_1(x)$, $v_1(x)$ are polynomials, it is necessary and sufficient that $h(x)$ should be divisible by $d(x)$. If here the degree of $h(x)$ is less than the sum of the degrees of $f(x)$ and $g(x)$, then the polynomials $u_1(x)$ and $v_1(x)$ can be chosen in such a way that the degree of $u_1(x)$ will be less than the degree of $g(x)$, and the degree of $v_1(x)$ less than the degree of $f(x)$. In this case $u_1(x)$ and $v_1(x)$ can be found if we write them with indeterminate coefficients and then equate the coefficients of similar powers of x on the left- and right-hand sides of equation (3.3) ($u_1(x)$ and $v_1(x)$ here cannot be unique). If, however, the degree of $h(x)$ is greater than or equal to the sum of the powers of $f(x)$ and $g(x)$, where $h(x) = d(x)k(x)$, then we must first represent $d(x)$ in the form (3.1), and then set $u_1(x) = u(x)k(x)$, $v_1(x) = v(x)k(x)$.

EXAMPLE 4.

$$f(x) = x^3 - x^2 - x + 1, \quad g(x) = x^2 + x - 2, \quad h(x) = x^4 - 1.$$

Here $d(x) = x - 1$, $h(x) = d(x)k(x)$, where $k(x) = x^3 + x^2 + x + 1$. Because the degree of $h(x)$ is less than the sum of the degrees of $f(x)$ and $g(x)$, we write the equation

$$x^4 - 1 = (x^3 - x^2 - x + 1)(ax + b) + (x^2 + x - 2)(a_1x^2 + b_1x + c_1).$$

Equating coefficients of similar powers of x:

$$\begin{cases} a + a_1 = 1, \\ -a + b + a_1 + b_1 = 0, \\ -a - b - 2a_1 + b_1 + c_1 = 0, \\ a - b - 2b_1 + c_1 = 0, \\ b - 2c_1 = -1. \end{cases}$$

The system of equations obtained has an uncountable number of solutions. Any of them gives coefficients of polynomials $u_1(x), v_1(x)$ which satisfy equation (3.3). For instance, we can take

$$u_1(x) = x + \frac{1}{3}, \quad v_1(x) = \frac{2}{3}(x + 1).$$

8. Eliminating irrationals in the denominator ([23], pp. 227–8).

The representation of the g.c.d. of the polynomials $f(x)$ and $g(x)$ in the form (3.1) is used in the solution of the following problem.

We have an expression of the form $m(\alpha)/g(\alpha)$, where $m(x)$ and $g(x)$ are polynomials with rational coefficients, and α is a root of the polynomial $f(x)$ also with rational coefficients, which is mutually prime to $g(x)$ (subsection 6). It is required to find a polynomial $N(x)$ with rational coefficients such that $m(\alpha)/g(\alpha) = N(\alpha)$. To solve the problem we represent $(f(x), g(x)) = 1$ in the form (3.1) and then we take the polynomial $m(x)v(x)$ as $N(x)$ (or, better, the remainder after division of this polynomial by $f(x)$).

EXAMPLE. To eliminate the irrationals in the denominator of the expression

$$\frac{10\sqrt[3]{(2)} - 10}{\sqrt[3]{(4)} - 2\sqrt[3]{(2)} + 2}.$$

Here $\alpha = \sqrt[3]{2}$, $m(x) = 10x - 10$, $g(x) = x^2 - 2x + 2$, $f(x) = x^3 - 2$, $(f(x), g(x)) = 1 = f(x)(-0{\cdot}1x - 0{\cdot}1) + g(x)(0{\cdot}1x^2 + 0{\cdot}3x + 0{\cdot}4)$, i.e. $v(x) = 0{\cdot}1x^2 + 0{\cdot}3x + 0{\cdot}4$, $m(x)v(x) = x^3 + 2x^2 + x - 4 = (x^3 - 2) \cdot 1 - 2x^2 + x - 2$. We set $N(x) = 2x^2 + x - 2$ and we obtain

$$\frac{10\sqrt[3]{(2)} - 10}{\sqrt[3]{(4)} - 2\sqrt[3]{(2)} + 2} = 2\sqrt[3]{4} + \sqrt[3]{2} - 2.$$

For another method of eliminating irrationals in the denominator, by using symmetric polynomials, see [23], p. 227.

9. The greatest common divisor of several polynomials.

If we are given several polynomials $f_1(x), \ldots, f_s(x)$, then their *greatest common divisor* is the polynomial $D(x)$, which divides all the given polynomials and which is itself divisible by any common divisor of the polynomials $f_1(x), \ldots, f_s(x)$. In order to find $D(x)$ it is necessary first to find the g.c.d. $d_1(x)$ of the polynomials $f_1(x)$ and $f_2(x)$, then the g.c.d. $d_2(x)$ of the polynomials $d_1(x)$ and $f_3(x)$, the g.c.d. $d_3(x)$ of the polynomials $d_2(x)$ and $f_4(x)$ and so on until $d_{s-1}(x) = (d_{s-2}(x), f_s(x)) = D(x)$.

EXAMPLE.

$$f_1(x) = x^2 - 2x + 2, \quad f_2(x) = x^2 + 2ix - 2, \quad f_3(x) = 2x^3 - 3x^2 - 2x + 1,$$
$$f_4(x) = x^3 - 2x^2 + x + 1.$$

Here $d_1(x) = x - (1 - i)$, $d_2(x) = 1$, $d_3(x) = D(x) = 1$.

10. The roots of a polynomial. The separation of multiple roots.

If when $x = c$ the polynomial $f(x)$ takes the value $f(c) = 0$, then the number c is called a *root* of this polynomial. Every polynomial of non-zero degree with any numerical coefficients has at least one root, real or complex (*the fundamental theorem of the algebra of complex numbers*). The number c is a root of the polynomial $f(x)$ if and only if $f(x)$ is divisible without remainder by $x - c$. If here $f(x)$ is divisible without remainder by $(x - c)^k$ $(k \geqslant 1)$, but is not divisible by $(x - c)^{k+1}$, then c is called a *k-tuple root* (or a *root of multiplicity k*) of the polynomial $f(x)$. A root of multiplicity $k = 1$ is called a *simple root* of the polynomial. If every root is reckoned as many times as its multiplicity, then we have that every polynomial of degree $n (n \geqslant 1)$ with any numerical coefficients has n roots.

In order to verify whether the number c is a root of the polynomial $f(x)$ and of what multiplicity, we can use Horner's scheme (subsection 3), where first $f(x)$ is divided by $x - c$, then, if the remainder is equal to zero, the quotient obtained is divided again by $x - c$ and so on until we obtain a non-zero remainder.

EXAMPLE 1.

$$f(x) = x^4 - 7x^3 + 9x^2 + 8x + 16, \quad c = 4.$$

We form Horner's scheme:

	1	−7	9	8	16
4	1	−3	−3	−4	0
	1	1	1	0	
	1	5	21		

$c = 4$ is a root of multiplicity 2.

The number c is a k-tuple root of the polynomial $f(x)$ if and only if the polynomial $f(x)$ and its derivatives† up to the order $k - 1$ inclusive vanish when $x = c$, but $f^{(k)}(c) \neq 0$. A root of multiplicity $k > 1$ of a polynomial is a root of multiplicity $k - 1$ of its first derivative.

† See footnote on p. 155.

If we divide the polynomial $f(x)$ by the g.c.d. $d(x)$ of this polynomial and its derivative $f'(x)$, then we obtain a polynomial $f(x)/d(x) = \varphi(x)$, which has the same roots as the polynomial $f(x)$, but only of the first multiplicity. This enables us to find all the roots of the polynomial $f(x)$ by finding all the roots of the polynomial $\varphi(x)$, whose degree is less than the degree of $f(x)$, provided that $d(x)$ is a polynomial of non-zero degree.

EXAMPLE 2.

$$f(x) = x^5 - 5x^4 - 5x^3 + 45x^2 - 108;$$

$$f'(x) = 5x^4 - 20x^3 - 15x^2 + 99x; \quad d(x) = x^3 - 4x^2 - 3x + 18;$$

$$\varphi(x) = \frac{f(x)}{d(x)} = x^2 - x - 6.$$

The roots of the polynomial $\varphi(x)$ are $\alpha_1 = -2$, $\alpha_2 = 3$. The same numbers (and only they) are roots of $f(x)$; their multiplicity is determined by means of Horner's scheme (p. 185).

We mention one further way sometimes used to find all the roots of a polynomial (and at the same time to determine their multiplicity). Let $d_1(x)$ be the g.c.d. of the polynomial $f(x)$ and its derivative $f'(x)$; $d_2(x)$ the g.c.d. of the polynomial $d_1(x)$ and its derivative $d_1'(x)$; $d_3(x)$ the g.c.d. of $d_2(x)$ and $d_2'(x)$ and so on, until we obtain a polynomial $d_s(x)$ of zero degree. We set

$$v_1(x) = \frac{f(x)}{d_1(x)}, \quad v_2(x) = \frac{d_1(x)}{d_2(x)}, \ldots, \quad v_s(x) = \frac{d_{s-1}(x)}{d_s(x)}$$

and then

$$F_1(x) = \frac{v_1(x)}{v_2(x)}, \quad F_2(x) = \frac{v_2(x)}{v_3(v)}, \ldots, \quad F_s(x) = v_s(x).$$

All the roots of the polynomials $F_1(x)$, $F_2(x)$, ..., $F_s(x)$ are simple, and the roots of $F_1(x)$ are all the simple roots of the polynomial $f(x)$ (and nothing else is), the roots of $F_2(x)$ are all the double roots of $f(x)$ and so on, the roots of $F_s(x)$ are the s-tuple roots of $f(x)$; the polynomial $f(x)$ cannot have roots above the sth multiplicity ([23], pp. 101–2; [13], p. 165).

EXAMPLE 3.

$$f(x) = x^5 + 6x^4 + 13x^3 + 14x^2 + 12x + 8.$$

Here

$$d_1(x) = x^2 + 4x + 4, \quad d_2(x) = x + 2, \quad d_3(x) = 1;$$

$$v_1(x) = x^3 + 2x^2 + x + 2, \quad v_2(x) = x + 2, \quad v_3(x) = x + 2;$$

$$F_1(x) = x^2 + 1, \quad F_2(x) = 1, \quad F_3(x) = x + 2.$$

The polynomial $f(x)$ has the simple roots $\pm i$ and the 3-tuple root 2.

11. Vieta's formulae.

The roots $\alpha_1, \ldots, \alpha_n$ of the polynomial

$$f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n \quad (a_0 \neq 0)$$

are connected with its coefficients by the formulae (Vieta's formulae):

$$\alpha_1 + \alpha_2 + \ldots + \alpha_n = -\frac{a_1}{a_0},$$

$$\alpha_1\alpha_2 + \alpha_1\alpha_3 + \ldots + \alpha_1\alpha_n + \alpha_2\alpha_3 + \ldots + \alpha_{n-1}\alpha_n = \frac{a_2}{a_0},$$

$$\alpha_1\alpha_2\alpha_3 + \alpha_1\alpha_2\alpha_4 + \ldots + \alpha_{n-2}\alpha_{n-1}\alpha_n = -\frac{a_3}{a_0},$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\alpha_1\alpha_2 \ldots \alpha_{n-1} + \alpha_1\alpha_2 \ldots \alpha_{n-2}\alpha_n + \ldots + \alpha_2\alpha_3 \ldots \alpha_n = (-1)^{n-1}\frac{a_{n-1}}{a_0},$$

$$\alpha_1\alpha_2 \ldots \alpha_n = (-1)^n\frac{a_n}{a_0},$$

i.e. the sum of the products of roots in all the possible combinations containing k roots is equal to $(-1)^k\, a_k/a_0$ $(k = 1, 2, \ldots, n)$. Using Vieta's formulae, it is easy to form a polynomial, given its roots in advance.

EXAMPLE. Let the roots of the polynomial be $\alpha_1 = 5$, $\alpha_2 = i$, $\alpha_3 = i$. A polynomial of the third degree, having these roots, will be, for instance, the polynomial with the coefficients $a_0 = 1$, $a_1 = -(5 + i + i) = -5 - 2i$, $a_2 = 5i + 5i + ii = -1 + 10i$, $a_3 = -(5ii) = 5$, i.e. $f(x) = x^3 + (-5 - 2i)x^2 + (-1 + 10i)x + 5$.

12. Common roots of two polynomials. Resultant. Discriminant.

Two polynomials $f(x)$ and $g(x)$ have common roots if and only if their g.c.d. $d(x)$ is not a constant. In this case the common roots of $f(x)$ and $g(x)$ are precisely all the roots of $d(x)$.

EXAMPLE 1.

$$f(x) = 2x^5 + x^4 + 3x^3 + x - 1, \quad g(x) = 2x^4 - 5x^3 + 4x^2 - 5x + 2.$$

Here $d(x) = 2x^3 - x^2 + 2x - 1 = (2x - 1)(x^2 + 1)$, the common roots of $f(x)$ and $g(x)$ are the numbers $1/2$, i, $-i$.

We mention one further method enabling us to establish the existence of common roots of the polynomials

$$f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$$

and

$$g(x) = b_0x^s + b_1x^{s-1} + \ldots + b_{s-1}x + b_s$$

($a_0 \neq 0$, $b_0 \neq 0$). We form the *resultant* of the given polynomials; this is what we call the determinant†

$$R(f, g) = \begin{vmatrix} a_0 & a_1 & \dots & a_n & 0 & \dots & 0 \\ 0 & a_0 & a_1 & \dots & a_n & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & a_0 & a_1 & \dots & a_n \\ b_0 & b_1 & \dots & b_s & 0 & \dots & 0 \\ 0 & b_0 & b_1 & \dots & b_s & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & b_0 & b_1 & \dots & b_s \end{vmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} s \text{ rows}, \\ \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} n \text{ rows}; \end{matrix}$$

it is equal to zero if and only if $f(x)$ and $g(x)$ have at least one common root.

EXAMPLE 2.

$$f(x) = x^3 - 3x^2 + 1, \quad g(x) = 2x^4 - 5x^2 - x - 1.$$

Here

$$R(f,g) = \begin{vmatrix} 1 & -3 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -3 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & -3 & 0 & 1 \\ 2 & 0 & -5 & -1 & -1 & 0 & 0 \\ 0 & 2 & 0 & -5 & -1 & -1 & 0 \\ 0 & 0 & 2 & 0 & -5 & -1 & -1 \end{vmatrix} = 543 \neq 0,$$

$f(x)$ and $g(x)$ have no common roots.

Other expressions for the resultant of the polynomials $f(x)$ and $g(x)$:

$$R(f, g) = a_0^s b_0^n \prod_{l=1}^{n} \prod_{j=1}^{s} (\alpha_l - \beta_j)$$

(α_i, β_j are the roots of the polynomials $f(x)$ and $g(x)$ respectively);

$$R(f, g) = a_0^s \prod_{i=1}^{n} g(\alpha_i);$$

$$R(f, g) = (-1)^{ns} R(g, f) = (-1)^{ns} b_0^n \prod_{j=1}^{s} f(\beta_j).$$

† A determinant of the $(n+s)$th order (see Chapter I, section 1).

The resultant of polynomials is used in the solution of systems of algebraic equations in several unknowns (section 5, subsection 4).

Discriminant of the polynomial $f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$, having as its roots the numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$, is the product

$$D(f) = a_0^{2n-2} \prod_{n \geqslant i > j \geqslant 1} (\alpha_i - \alpha_j)^2.$$

The discriminant is equal to zero if and only if some of the roots of the polynomial are equal (i.e. the polynomial has at least one multiple root (subsection 10)). The discriminant is connected to the resultant of the polynomial $f(x)$ and its derivative $f'(x)$ by the formula

$$R(f, f') = (-1)^{\frac{n(n-1)}{2}} a_0 D(f),$$

which enables us to express the discriminant of a polynomial in terms of its coefficients.

EXAMPLE 3.

$$f(x) = x^3 - 3x + 7.$$

Here $f'(x) = 3x^2 - 3$,

$$(-1)^3 D(f) = R(f, f') = \begin{vmatrix} 1 & 0 & -3 & 7 & 0 \\ 0 & 1 & 0 & -3 & 7 \\ 3 & 0 & -3 & 0 & 0 \\ 0 & 3 & 0 & -3 & 0 \\ 0 & 0 & 3 & 0 & -3 \end{vmatrix} = 1215.$$

$D(f) = -1215$; $f(x)$ does not have multiple roots.

We can express $D(f)$ in terms of the coefficients of the polynomial $f(x)$ in yet another way, by using the fact that $D(f)$ is a symmetric polynomial in the roots of $f(x)$ (section 5, subsection 2).

The expression for the discriminant in terms of the power sums (section 5, subsection 3) is:

$$D(f) = a_0^{2n-2} \begin{vmatrix} n & s_1 & s_2 & \ldots & s_{n-1} \\ s_1 & s_2 & s_3 & \ldots & s_n \\ s_2 & s_3 & s_4 & \ldots & s_{n+1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ s_{n-1} & s_n & s_{n+1} & \ldots & s_{2n-2} \end{vmatrix}$$

(s_i is the sum of the ith powers of the roots $\alpha_1, \alpha_2, \ldots, \alpha_n$ of the polynomial $f(x)$).

EXAMPLE 4.

$$f(x) = x^3 + 3x^2 - 2.$$

According to sections 5 (the end of subsections 2 and 3), we find $s_1 = -3$, $s_2 = 9$, $s_3 = -21$, $s_4 = 57$. Hence

$$D(f) = \begin{vmatrix} 3 & s_1 & s_2 \\ s_1 & s_2 & s_3 \\ s_2 & s_3 & s_4 \end{vmatrix} = \begin{vmatrix} 3 & -3 & 9 \\ -3 & 9 & -21 \\ 9 & -21 & 57 \end{vmatrix} = 108.$$

If $f(x)$ is a polynomial with real coefficients, not having multiple roots, then $D(f) > 0$, if the number of pairs of complex conjugate roots of $f(x)$ is even, and $D(f) < 0$, if this number is odd.

EXAMPLE 5. For the polynomial

$$f(x) = x^4 - x^3 + x^2 + 2x + 2$$

$D(f) = 4360 > 0$, i.e. either all four of the roots of the polynomial $f(x)$ are complex (two pairs of complex conjugate roots) or $f(x)$ in general does not have complex roots (the number of pairs of complex conjugate roots of $f(x)$ is equal to zero).

EXAMPLE 6. $f(x) = 2x^3 + 9x^2 + 9x + 7$.

Here $D(f) = -9126 < 0$, $f(x)$ has one pair of complex conjugate roots, i.e. it has two complex and one real root.

2. The Interpolation Formulae of Lagrange and Newton

If $b_0, b_1, \ldots, b_n$ are numbers which are all different, and $c_0, c_1, \ldots, c_n$ are any numbers not all equal to zero, then there exists precisely one polynomial $f(x)$ of degree not greater than n with the properties $f(b_0) = c_0$, $f(b_1) = c_1, \ldots, f(b_n) = c_n$.† This polynomial can be found in one of the following three ways:

(1) *We form the system of equations*:

$$\begin{aligned} b_0^n x_0 + b_0^{n-1} x_1 + \ldots + b_0 x_{n-1} + x_n &= c_0, \\ b_1^n x_0 + b_1^{n-1} x_1 + \ldots + b_1 x_{n-1} + x_n &= c_1, \\ \cdots\cdots\cdots\quad \cdots\cdots\cdots & \\ b_n^n x_0 + b_n^{n-1} x_1 + \ldots + b_n x_{n-1} + x_n &= c_n. \end{aligned}$$

† When $c_0 = c_1 = \ldots = c_n = 0$ every polynomial with the properties $f(b_i) = c_i$ $(i = 0, 1, , 2 \ldots, n)$ either has degree greater than n or is identically equal to zero.

If $a_0, a_1, \ldots, a_{n-1}, a_n$ is a solution of this system of equations (Chapter I, section 2, subsections 1–3), then the desired polynomial is:

$$f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n.$$

EXAMPLE 1. $b_0 = -1$, $b_1 = 1$, $b_2 = 2$, $b_3 = 3$; $c_0 = 4$, $c_1 = 2$, $c_2 = -2$, $c_3 = 0$.

We form the system of equations:

$$\begin{aligned} -x_0 + x_1 - x_2 + x_3 &= 4, \\ x_0 + x_1 + x_2 + x_3 &= 2, \\ 8x_0 + 4x_1 + 2x_2 + x_3 &= -2, \\ 27x_0 + 9x_1 + 3x_2 + x_3 &= 0. \end{aligned}$$

Solving this system, we find $a_0 = 1$, $a_1 = -3$, $a_2 = -2$, $a_3 = 6$. The desired polynomial is $f(x) = x^3 - 3x^2 - 2x + 6$.

(2) *Lagrange's formula*:

$$f(x) = \sum_{k=0}^{n} c_k \frac{(x-b_0)\ldots(x-b_{k-1})(x-b_{k+1})\ldots(x-b_n)}{(b_k-b_0)\ldots(b_k-b_{k-1})(b_k-b_{k+1})\ldots(b_k-b_n)},$$

or, in another notation,

$$f(x) = c_0\frac{g_0(x)}{g'(b_0)} + c_1\frac{g_1(x)}{g'(b_1)} + \ldots + c_n\frac{g_n(x)}{g'(b_n)}, \tag{3.4}$$

where

$$g(x) = (x-b_0)(x-b_1)\ldots(x-b_n),$$

and $g_k(x)$, $k = 0, 1, \ldots, n$, is the product of the same brackets, with the omission of the bracket $(x - b_k)$, $g'(x)$ is the derivative of the polynomial $g(x)$ ([23], p. 110).

EXAMPLE 2. $b_0 = 1$, $b_1 = 3$, $b_2 = 4$; $c_0 = 2$, $c_1 = -2$, $c_2 = -1$. The desired polynomial is:

$$f(x) = 2\frac{(x-3)(x-4)}{(1-3)(4-4)} - 2\frac{(x-1)(x-4)}{(3-1)(3-4)} - \frac{(x-1)(x-3)}{(4-1)(4-3)}$$
$$= x^2 - 6x + 7.$$

Lagrange's formula is convenient to use if we do not need to determine the coefficients of the polynomial $f(x)$ with the desired properties but only its value for a certain value $x = b_{n+1}$.

EXAMPLE 3. Knowing that $f(x)$ is a polynomial of degree at most 3 and that $f(-2) = 0$, $f(1) = 5$, $f(3) = -1$, $f(5) = -1$, to find $f(6)$.

By virtue of Lagrange's formula we have

$$f(6) = 5\frac{(6+2)(6-3)(6-5)}{(1+2)(1-3)(1-5)} - \frac{(6+2)(6-1)(6-5)}{(3+2)(3-1)(3-5)} - \frac{(6+2)(6-1)(6-3)}{(5+2)(5-1)(5-3)} = 4\frac{6}{7}.$$

Lagrange's formula also gives the expansion of a fraction into partial fractions in the case when the denominator does not have multiple roots and can be decomposed over the given field into linear factors (section 3, subsection 9). If $b_0, b_1, \ldots, b_n$ are the roots of the denominator of the fraction $f(x)/g(x)$, where $g(x)$ is a polynomial of the $(n+1)$th degree, $f(x)$ is a polynomial of degree $\leqslant n$, and if $f(b_0) = c_0, f(b_1) = c_1, \ldots, f(b_n) = c_n$, then we express $f(x)$ by means of the formula (3.4) and we obtain

$$\frac{f(x)}{g(x)} = \frac{c_0}{g'(b_0)(x-b_0)} + \frac{c_1}{g'(b_1)(x-b_1)} + \ldots + \frac{c_n}{g'(b_n)(x-b_n)}.$$

EXAMPLE 4. To decompose the fraction $\Phi(x) = (x+1)/(x^2-x-6)$ into partial fractions. Here

$$b_0 = -2,\ b_1 = 3;\ c_0 = -1,\ c_1 = 4;\ g'(b_0) = -5,\ g'(b_1) = 5.$$

$$\Phi(x) = \frac{1}{5(x+2)} + \frac{4}{5(x-3)}.$$

(3) *Newton's formula* ([23], p. 111):

$$f(x) = \lambda_0 + \lambda_1(x-b_0) + \lambda_2(x-b_0)(x-b_1) + \lambda_n(x-b_0)\ldots(x-b_{n-1}),$$

where we find λ_0 by setting $x = b_0$ on both sides of the equation ($\lambda_0 = f(b_0) = c_0$), then we find λ_1 by setting $x = b_1$ on both sides of the equation, and so on up to λ_n, which we find by setting $x = b_n$ on both sides of the equation.

EXAMPLE 5.

$$b_0 = 1, b_1 = -2, b_2 = 3; c_0 = 4,\ c_1 = 7,\ c_2 = 12.$$

$$f(x) = 4 + \lambda_1(x-1) + \lambda_2(x-1)(x+2).$$

Setting $x = -2$, we obtain $7 = 4 + \lambda_1(-3)$, $\lambda_1 = -1$. We take $x = 3$: $12 = 4 - (3-1) + \lambda_2(3-1)(3+2)$, $10 = 10\lambda_2$, $\lambda_2 = 1$. The desired polynomial is $f(x) = 4 - (x-1) + (x-1)(x+2) = x^2 + 3$.

If $b_0, b_1, \ldots, b_n$ are consecutive integers, then it is convenient to carry out the calculation by means of a scheme where in the first column we write the numbers $c_0, c_1, \ldots, c_n$, in the second their differences $\Delta c_0 = c_1 - c_0, \ldots, \Delta c_{n-1} = c_n - c_{n-1}$, in

the third column the differences of the numbers in the second column $\Delta^2 c_0 = \Delta c_1 - \Delta c_0, \ldots, \Delta^2 c_{n-2} = \Delta c_{n-1} - \Delta c_{n-2}$ and so on ([5], Chapter I, p. 97):

$$\begin{array}{lllll} c_0 & & & & \\ & \Delta c_0 & & & \\ c_1 & & \Delta^2 c_0 & & \\ & \Delta c_1 & \vdots & \vdots & \Delta^n c_0 \\ c_2 & \vdots & & & \\ \vdots & \Delta c_{n-1} & & & \\ c_n & & & & \end{array}$$

Then $\lambda_0 = c_0$, $\lambda_1 = \Delta c_0$, $\lambda_2 = \Delta^2 c_0/2!, \ldots, \lambda_n = \Delta^n c_0/n!$.

EXAMPLE 6.

$$b_0 = -2,\ b_1 = -1,\ b_2 = 0,\ b_3 = 1,\ b_4 = 2;\ c_0 = -8,$$
$$c_1 = -3, c_2 = 4, c_3 = 13, c_4 = 24.$$

We write out the scheme:

$$\begin{array}{lllll} -8 & & & & \\ & 5 & & & \\ -3 & & 2 & & \\ & 7 & & 0 & \\ 4 & & 2 & & 0 \\ & 9 & & 0 & \\ 13 & & 2 & & \\ & 11 & & & \\ 24 & & & & \end{array}$$

The desired polynomial is

$$f(x) = -8 + 5(x+2) + (x+2)(x+1) = x^2 + 8x + 4.$$

If we have already constructed the polynomial $f_1(x)$ of degree not more than $s (s < n)$, which takes the given values $c_0, c_1, \ldots, c_s$ when $x = b_0, b_1, \ldots, b_s$, then the polynomial $f(x)$ of degree not more than n with the properties $f(b_i) = c_i$, $i = 0, 1, \ldots, n$, is constructed by means of the formula

$$f(x) = f_1(x) + \lambda_{s+1}(x - b_0)\ldots(x - b_s) + \ldots$$
$$+ \lambda_n(x - b_0)\ldots(x - b_{n-1}),$$

where we find λ_{s+1} by setting $x = b_{s+1}$, and so on ([23], p. 112).

EXAMPLE 7.

$$b_0 = 1,\ b_1 = -2,\ b_2 = 3,\ b_3 = 0,\ b_4 = 4;\ c_0 = 4,\ c_1 = 7,$$
$$c_2 = 12,\ c_3 = 3,\ c_4 = -5.$$

The polynomial $f_1(x) = x^2 + 3$ possesses the properties $f_1(1) = 4$, $f_1(-2) = 7$, $f_1(3) = 12$, $f_1(0) = 3$. The desired polynomial is

$$f(x) = x^2 + 3 + \lambda_4(x-1)(x+2)(x-3)x.$$

We take $x = 4$: $-5 = 19 + \lambda_4(4-1)(4+2)(4-3)\times 4$, $\lambda_4 = -\frac{1}{3}$.

Hence,

$$f(x) = -\frac{1}{3}x^4 + \frac{2}{3}x^3 + \frac{8}{3}x^2 - 2x + 3.$$

The general form of all the polynomials which take the values $c_0, c_1, \ldots, c_n$ respectively when $x = b_0, b_1, \ldots, b_n$ is

$$F(x) = f(x) + \varphi(x)(x-b_0)\ldots(x-b_n),$$

where $f(x)$ is a polynomial of degree at most n with the properties $f(b_0) = c_0, \ldots, f(b_n) = c_n$, and $\varphi(x)$ is an arbitrary polynomial ([13], p. 159).

3. Location of the Roots of Polynomials. The Decomposition of Polynomials into Factors

1. Solution of equations of the second degree.

The roots of the quadratic equation $ax^2 + bx + c = 0$ with any real or complex coefficients are found by means of the formula

$$x = \frac{-b \pm \sqrt{(b^2 - 4ac)}}{2a}.$$

For the roots of the "reduced" quadratic equation $x^2 + px + q = 0$ we have the formula

$$x = -\frac{p}{2} \pm \sqrt{\left(\frac{p^2}{4} - q\right)}.$$

EXAMPLE 1. To solve the equation $(6+3i)x^2 + (5+4i)x + (1+i) = 0$. We obtain

$$x = \frac{-5-4i \pm \sqrt{(9+40i-12-36i)}}{12+6i} = \frac{-5-4i \pm \sqrt{(-3+4i)}}{12+6i}$$

$$= \frac{-5-4i \pm (1+2i)}{12+6i}, \quad x_1 = \frac{-4-2i}{12+6i} = -\frac{1}{3},$$

$$x_2 = \frac{-6-6i}{12+6i} = -\frac{3}{5} - \frac{1}{5}i.$$

EXAMPLE 2. To solve the equation $x^2 + ix - (4 - 2i) = 0$.
We have

$$x = -\frac{i}{2} \pm \sqrt{\left(-\frac{1}{4}\right) + 4 - 2i} = -\frac{i}{2} \pm \frac{4 - i}{2}$$

i.e. $x_1 = 2 - i$, $x_2 = -2$.

2. Solution of equations of the third degree.

In order to find all the roots of the equation of the third degree

$$a_0x^3 + a_1x^2 + a_2x + a_3 = 0,$$

it is convenient to divide both sides of this equation by a_0, giving an equation of the form

$$x^3 + b_1x^2 + b_2x + b_3 = 0, \tag{3.5}$$

which has the same roots as the original equation, and then to make the change of unknown $x = y - b_1/3$. It is easiest to make this change if we expand the polynomial $f(x) = x^3 + b_1x^2 + b_2x + b_3$ in powers of $x + b_1/3$ with the help of Horner's scheme (section 1, subsection 4) and then replace $x + b_1/3$ by y. As a result of this substitution we obtain an equation of the form

$$y^3 + py + q = 0. \tag{3.6}$$

The roots of this equation are found by means of the formula

$$y = \sqrt[3]{\left\{-\frac{q}{2} + \sqrt{\left[\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3\right]}\right\}} + \sqrt[3]{\left\{-\frac{q}{2} - \sqrt{\left[\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3\right]}\right\}} \tag{3.7}$$

(*Cardan's formula*). Applying this formula, it is necessary, for each of the three values of the root

$$\alpha = \sqrt[3]{\left\{-\frac{q}{2} + \sqrt{\left[\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3\right]}\right\}}, \tag{3.8}$$

to take the values of the root

$$\beta = \sqrt[3]{\left\{-\frac{q}{2} - \sqrt{\left[\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3\right]}\right\}}, \tag{3.9}$$

for which the condition $\alpha\beta = -p/3$ is satisfied (such a value for the root β always exists). If y_1, y_2, y_3 are the roots of the

equation (3.6), then $x_1 = y_1 - b_1/3$, $x_2 = y_2 - b_1/3$, $x_3 = y_3 - b_1/3$ are the roots of equation (3.5).

EXAMPLE 1. To solve the equation

$$x^3 + 3x^2 - 6x + 30 = 0.$$

We expand the polynomial on the left-hand side of the equation in powers of $x + 1$ and we set $x + 1 = y$. We obtain the equation $y^3 - 9y + 28 = 0$. Its roots are found by means of the formula

$$y = \sqrt[3]{[-14 + \sqrt{(196 - 27)}]} + \sqrt[3]{[-14 - \sqrt{(196 - 27)}]}.$$

The values of the root $\alpha = \sqrt[3]{(-14 + \sqrt{169})} = \sqrt[3]{-1}$ are the numbers

$$\alpha_1 = -1, \alpha_2 = \frac{1}{2} + \frac{\sqrt{3}}{2} i, \quad \alpha_3 = \frac{1}{2} - \frac{\sqrt{3}}{2} i.$$

The values of the other root corresponding to them are

$$\beta_1 = -\frac{-9}{3\alpha_1} = -3, \quad \beta_2 = -\frac{-9}{3\alpha_2} = \frac{3}{2} - \frac{3\sqrt{3}}{2} i,$$

$$\beta_3 = -\frac{-9}{3\alpha_3} = \frac{3}{2} + \frac{3\sqrt{3}}{2} i.$$

Hence $y_1 = \alpha_1 + \beta_1 = -4$, $y_2 = \alpha_2 + \beta_2 = 2 - i\sqrt{3}$, $y_3 = \alpha_3 + \beta_3 = 2 + i\sqrt{3}$. The roots of the original equation are $x_1 = -5$, $x_2 = 1 - i\sqrt{3}$, $x_3 = 1 + i\sqrt{3}$.

The roots of equation (3.6) can also be found by means of the formulae

$$\left.\begin{aligned} y_1 &= \alpha_1 + \beta_1, \\ y_2 &= -\frac{1}{2}(\alpha_1 + \beta_1) + i\frac{\sqrt{3}}{2}(\alpha_1 - \beta_1), \\ y_3 &= -\frac{1}{2}(\alpha_1 + \beta_1) - i\frac{\sqrt{3}}{2}(\alpha_1 - \beta_1), \end{aligned}\right\} \tag{3.10}$$

where as α_1 we take any of the values of the roots of (3.8), and as β_1 the value of the roots of (3.9) for which $\alpha_1 \beta_1 = -p/3$.

EXAMPLE 2. To solve the equation $x^3 - 9x^2 + 21x - 5 = 0$.

We make the substitution $x = y + 3$ and we obtain the equation $y^3 - 6y + 4 = 0$. For this one of the values of the root $\alpha = \sqrt[3]{[-2 + \sqrt{(4 - 8)}]} = \sqrt[3]{(-2 + 2i)}$ is $\alpha_1 = 1 + i$. The value of the other root corresponding to it is $\beta_1 = -6/[3(1 + i)] = 1 - i$. By formulae (3.10) we find: $y_1 = 2$, $y_2 = -1 - \sqrt{3}$, $y_3 = -1 + \sqrt{3}$. The roots of our given equation are $x_1 = 5$, $x_2 = 2 - \sqrt{3}$, $x_3 = 2 + \sqrt{3}$.

If in equation (3.6) $p = 0$, then the roots of this equation are all the values of $\sqrt[3]{-q}$; if $q = 0$, then the roots of equation

(3.6) are $y_1 = 0$, $y_{2,3} = \pm\sqrt{-p}$. If $p \neq 0$ and $q \neq 0$, but $(q/2)^2 + (p/3)^3 = 0$ then the roots of equation (3.6) are found thus:

$$y_1 = \frac{3q}{p}, \quad y_2 = y_3 = -\frac{3q}{2p} \tag{3.11}$$

([17], p. 215).

EXAMPLE 3. To solve the equation $x^3 + 3x - 2i = 0$.

Here there is no term in x^2, so that this equation is already of the form (3.6); $p = 3$, $q = -2i$, $(q/2)^2 + (p/3)^2 = 0$. The formulae (3.11) give $x_1 = -2i$, $x_2 = x_3 = i$.

If all the coefficients of the equation (3.6) are real numbers, then its roots can be found without explicitly using Cardan's formula (3.7) in the following way ([21], pp. 452–4):

First case. Let $D = (q/2)^2 + (q/3)^2 < 0$.† Then all three of the roots of equation (3.6) are real and are found from the formulae

$$y_k = 2\sqrt{\left(-\frac{p}{3}\right)} \cos\frac{\varphi + 2k\pi}{3}, \ddagger \tag{3.12}$$

where $k = 1, 2, 3$, and φ is the smallest positive angle satisfying the condition $\cos\varphi = -\frac{q}{2}\sqrt{[-(3/p)^3]}$.

EXAMPLE 4.

$$y^3 - 3y + 1 = 0. \text{ Here } D = \tfrac{1}{4} - 1 < 0.$$

We apply the formulae (3.12) $\cos\varphi = -\frac{1}{2}\sqrt{1} = -\frac{1}{2}$, $\varphi = 120°$; $y_k = 2\cos(120° + 2k\pi)/3$, $k = 1, 2, 3$, i.e. $y_1 = 2\cos 160° \approx -1{\cdot}880$, $y_2 = 2\cos 280° \approx 0{\cdot}348$, $y_3 = 2\cos 40° \approx 1{\cdot}532$.

Second case. Let $D > 0$. Then equation (3.6) has one real and two complex conjugate roots. If $p < 0$, then these roots are found thus:

$$y_1 = -\frac{2\sqrt{-\frac{p}{3}}}{\sin 2\varphi}, \quad y_{2,3} = \frac{\sqrt{-\frac{p}{3}}}{\sin 2\varphi} \pm i\sqrt{(-p)}\cot 2\varphi, \tag{3.13}$$

where

$$\tan\varphi = \sqrt[3]{\left(\tan\frac{\omega}{2}\right)}, \quad \sin\omega = \frac{2}{q}\sqrt{\left[-\left(\frac{p}{3}\right)^3\right]}.$$

† The number $-108\,D$ is the discriminant of the polynomial on the left-hand side of the equation (3.6) (section 1, subsection 12).

‡ Here and in the following formulae we must take the positive values of the square root.

If $p > 0$, then we have the formulae

$$y_1 = -2\sqrt{\left(\frac{p}{2}\right)}\cot 2\varphi, \quad y_{2,3} = \sqrt{\left(\frac{p}{3}\right)}\cot 2\varphi \pm \frac{i\sqrt{p}}{\sin 2\varphi}, \qquad (3.14)$$

where

$$\tan\varphi = \sqrt[3]{\left(\tan\frac{\omega}{2}\right)}, \quad \tan\omega = \frac{2}{q}\sqrt{\left[\left(\frac{p}{3}\right)^3\right]}.$$

EXAMPLE 5.

$$y^3 - 9y + 12 = 0.$$

Here $D = 36 - 27 > 0$, $p = -9 < 0$.
We apply formulae (3.13);

$\sin\omega = \frac{2}{12}\sqrt{3^3} = \frac{\sqrt{3}}{2}$, $\omega = 60°$; $\tan\varphi = \sqrt[3]{\tan 30°} \approx 0{\cdot}833$, $\varphi \approx 30°50'$;

$\sin 2\varphi \approx 0{\cdot}984$; $\cot 2\varphi \approx 0{\cdot}182$; $y_1 \approx -\frac{2\sqrt{3}}{0{\cdot}984} \approx -3{\cdot}520$, $y_{2,3} \approx \frac{\sqrt{3}}{0{\cdot}984}$ $\pm i \times 3 \times 0{\cdot}182 \approx 1{\cdot}760 \pm 0{\cdot}546i$.

EXAMPLE 6.

$$x^3 - 6x^2 + 17x - 25 = 0.$$

The substitution $y = x - 2$ gives the equation $y^3 + 5y - 7 = 0$, for which $D = 49/4 + 125/27 > 0$, $p = 5 > 0$. We apply formulae (3.14), and we perform the calculation with the help of logarithms:

$\tan\omega = \frac{2}{-7}\sqrt{\frac{125}{27}}$, $\tan(180° - \omega) = -\tan\omega = \frac{2}{7}\sqrt{\frac{125}{27}}$, $\log\tan$ $(180° - \omega) = 1{\cdot}7886$, $180° - \omega = 31°35'$, $\omega = 148°25'$; $\tan\varphi = \sqrt[3]{(\tan 74°13')}$, $\log\tan\varphi = 0{\cdot}1829$, $\varphi = 56°44'$; $y_1 = -2\sqrt{\left(\frac{5}{3}\right)}\cot 113°28' = 2\sqrt{\left(\frac{5}{3}\right)}$ $\cot 66°32'$, $\log y_1 = 0{\cdot}0496$, $y_1 = 1{\cdot}121$; $y_{2,3} = \sqrt{\left(\frac{5}{3}\right)}\cot 113°28'$ $\pm \frac{i\sqrt{5}}{\sin 113°28'} = -0{,}561 \pm \frac{i\sqrt{5}}{\sin 66°32'} = -0{\cdot}561 \pm 2{\cdot}438i$; $x_1 = 3{\cdot}121$, $x_{2,3} = 1{\cdot}439 \pm 2{\cdot}438i$.

In the case $D > 0$ the roots of equation (3.6) can also be found by means of the formulae ([4], p. 139):

when $p < 0$ $y_1 = -2r\cosh\varphi/3$, $y_{2,3} = r\cosh\varphi/3 \pm i\sqrt{3}r\sinh\varphi/3$, where $\cosh\varphi = q/2r^3$, $r = \pm\sqrt{-p/3}$, where the sign $+$ or $-$ is chosen depending on whether $q > 0$ or $q < 0$;

when $p > 0$ $y_1 = -2r\sinh\varphi/3$, $y_{2,3} = r\sinh\varphi/3 \pm i\sqrt{3}r\cosh\varphi/3$, where $\sinh\varphi = q/2r^3$, $r = \pm\sqrt{p/3}$, where the sign of the square root is again chosen the same as the sign of q.

EXAMPLE 7.

$$y^3 - 6y + 9 = 0.$$

Here $q^2/4 + p^3/27 = 81/4 - 8 > 0$, $p = -6 < 0$, $q = 9 > 0$, and we have: $r = \sqrt{2} \approx 1{\cdot}41$, $\cosh\varphi = \dfrac{9}{2 \times 2\sqrt{2}} \approx 1{\cdot}59$, $\varphi \approx 1{\cdot}05$; $y_1 \approx -2\sqrt{2}\cosh 0{\cdot}35 \approx -2{\cdot}99$, $y_{2,3} \approx \sqrt{2}\cosh 0{\cdot}35 \pm i\sqrt{3}\sqrt{2}\sinh 0{\cdot}35 \approx 1{\cdot}49 \pm 0{\cdot}86i$.

EXAMPLE 8.

$$y^3 + 9y - 26 = 0.$$

Here $q^2/4 + p^3/27 > 0$, $p = 9 > 0$, $q = -26 < 0$, and we have: $r = -\sqrt{3} \approx -1{\cdot}73$, $\sinh\varphi = \dfrac{-13}{-3\sqrt{3}} \approx 2{\cdot}51$, $\varphi \approx 1{\cdot}65$; $y_1 \approx 2\sqrt{3}\sinh 0{\cdot}55 \approx 2{\cdot}01$, $y_{2,3} \approx -\sqrt{3}\sinh 0{\cdot}55 \mp i\sqrt{3}\sqrt{3}\cosh 0{\cdot}55 \approx -1 \mp 3{\cdot}48i$.

Third case. Let $D = 0$. Then all the roots of the equation (3.6) are real, and two of them are equal to each other. The roots are calculated by means of the formulae (3.11) (when $p \neq 0$).

3. Solution of equations of the fourth degree.

In order to find all the roots of an equation of the fourth degree, we divide both sides by the coefficient of x^4 (which does not change the roots of the equation) and then in the equation

$$x^4 + b_1x^3 + b_2x^2 + b_3x + b_4 = 0 \tag{3.15}$$

we make the change of variables $x = y - b_1/4$ (using Horner's scheme, as in subsection 2). As a result we obtain the equation

$$y^4 + py^2 + qy + r = 0. \tag{3.16}$$

If y_1, y_2, y_3, y_4 are the roots of equation (3.16), then $x_i = y_i - b_1/4$ $(i = 1, 2, 3, 4)$ are the roots of equation (3.15).

The roots of equation (3.16) can be found in the following way.

(1) *Ferrari's method.* We form the cubic equation

$$z^3 + 2pz^2 + (p^2 - 4r)z - q^2 = 0 \tag{3.17}$$

(the *cubic resolvent* of the equation (3.16)) and take one of its roots z_0. Then we form two quadratic equations:

$$u^2 - \sqrt{z_0}u + \left(\frac{p + z_0}{2} + \frac{q}{2\sqrt{z_0}}\right) = 0, \tag{3.18}$$

$$u^2 + \sqrt{z_0}u + \left(\frac{p + z_0}{2} - \frac{q}{2\sqrt{z_0}}\right) = 0. \tag{3.19}$$

The roots of the equation (3.18) and (3.19) are the roots of the equation (3.16).

EXAMPLE 1. To solve the equation $x^4 - 4x^3 + 7x^2 - 2x - 5 = 0$.

The substitution $y = x - 1$ gives $y^4 + y^2 + 4y - 3 = 0$. The cubic resolvent of this equation is $z^3 + 2z^2 + 13z - 16 = 0$; one of its roots is $z_0 = 1$. Equations (3.18) and (3.19) are $u^2 - u + 3 = 0$, $u^2 + u - 1 = 0$; their roots are

$$y_{1,2} = \frac{1}{2} \pm \sqrt{\left(\frac{1}{4} - 3\right)} = \frac{1 + i\sqrt{11}}{2},$$

$$y_{3,4} = -\frac{1}{2} \pm \sqrt{\left(\frac{1}{4} + 1\right)} = \frac{-1 \pm \sqrt{5}}{2}.$$

The roots of our given equation are

$$x_{1,2} = \frac{3 \pm i\sqrt{11}}{2}, \quad x_{3,4} = \frac{1 \pm \sqrt{5}}{2}.$$

(2) *Euler's method.* ([23], pp. 204–5, [17], pp. 222–5, or [31], pp. 156–8). We form the cubic resolvent (3.17) and we find its roots z_0, z_1, z_2. We find the roots of equation (3.16) by means of the formulae

$$y_1 = \frac{1}{2}(\sqrt{z_0} + \sqrt{z_1} + \sqrt{z_2}),$$

$$y_2 = \frac{1}{2}(\sqrt{z_0} - \sqrt{z_1} - \sqrt{z_2}),$$

$$y_3 = \frac{1}{2}(-\sqrt{z_0} + \sqrt{z_1} - \sqrt{z_2}),$$

$$y_4 = \frac{1}{2}(-\sqrt{z_0} - \sqrt{z_1} + \sqrt{z_2}),$$

where the signs of $\sqrt{z_0}, \sqrt{z_1}, \sqrt{z_2}$ are chosen in such a way that the equation $\sqrt{z_0} \cdot \sqrt{z_1} \cdot \sqrt{z_2} = -q$ is satisfied.

EXAMPLE 2.

$$y^4 - 6y^2 + 8y - 3 = 0.$$

Here the cubic resolvent $z^3 - 12z^2 + 48z - 64 = 0$ has roots $z_0 = z_1 = z_2 = 4$; $\sqrt{z_i} = \pm 2$, $i = 0, 1, 2$. The roots of our given equation of the fourth degree are:

$$y_1 = \frac{1}{2}(2 + 2 - 2) = 1, \quad y_2 = \frac{1}{2}(2 - 2 + 2) = 1,$$

$$y_3 = \frac{1}{2}(-2 + 2 + 2) = 1, \quad y_4 = \frac{1}{2}(-2 - 2 - 2) = -3.$$

(3) If in equation (3.16) $q = 0$ (a *biquadratic* equation), then its roots are found by means of the formula

$$y = \pm \sqrt{\left[-\frac{p}{2} \pm \sqrt{\left(\frac{p^2}{4} - r \right)} \right]}.$$

EXAMPLE 3.

$$y^4 + 10y^2 + 9 = 0; \; y = \pm \sqrt{[-5 \pm \sqrt{(25 - 9)}]};$$

the roots of the equation are $y_{1,2} = \pm i$, $y_{3,4} = \pm 3i$.

4. Equations of degree higher than the fourth.

We say that the equation

$$a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n = 0$$

is *solvable by radicals* if its roots can be expressed in terms of its coefficients using the operations of addition, subtraction, multiplication, division and the extraction of roots. Any equation of degree $\leqslant 4$ can be solved by radicals, and for every fixed $n \leqslant 4$ we can give a general formula by which the roots of any equation of degree n can be expressed in terms of the coefficients of this equation. For $n > 4$ it is impossible to give such a formula (*the Ruffini–Abel theorem*) ([23], p. 392, or [17], p. 279). Moreover, for every $n > 4$ we can give an equation of degree n with integral coefficients such that its roots in general cannot be expressed in terms of its coefficients with the use of the operations mentioned above (for $n = 5$ such an equation will be, for instance, the equation $x^5 - 4x - 2 = 0$). But, in addition, for every n there also exists an equation of degree n which can be solved by radicals, for instance the *binomial equation* $x^n - a = 0$, the roots of which are all the values of the nth roots of the number a ([11], p. 126). The problem of the conditions under which a given equation is solvable by radicals was investigated by Galois in the thirties of the last century ([5], Chapter 1, p. 221).

5. Rational roots of polynomials with rational coefficients.

The polynomial $f(x)$ with rational coefficients has the same roots as the polynomial $g(x)$ with integral coefficients obtained from $f(x)$ by multiplying it by a common multiple of all the denominators of the coefficients of $f(x)$.

The rational roots of the polynomial

$$g(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n \tag{3.20}$$

with integral coefficients can only be irreducible fractions of the form p/q, in which p is a divisor of the number a_n, q is a divisor of the number a_0, and in fact only those of these fractions for which with any integral m, $g(m)$ is divisible by $p - mq$. If $a_0 = 1$, then all the rational roots of the polynomial (3.20) (if in general it has any) are necessarily integers.

In order to find all the rational roots of the polynomial (3.20), we can first multiply this polynomial by a_0^{n-1} and make the substitution $y = a_0 x$, as a result of which we obtain the polynomial

$$h(y) = y^n + b_1 y^{n-1} + \ldots + b_{n-1} y + b_n \tag{3.21}$$

again with integral coefficients. In order to find all the rational roots of $g(x)$, it is sufficient to find all the integral roots of $h(y)$ and divide them by a_0. To find the integral roots of $h(y)$ we must, for instance with the help of Horner's scheme, find which of the divisors of the number b_n are roots of $h(y)$ (section 1, subsection 10), and we need not consider those divisors p for which $h(1)$ is not divisible by $p - 1$ or $h(-1)$ is not divisible by $p + 1$; further we can take any integer $m \neq \pm 1$ and exclude from our considerations also those divisors p for which $h(m)$ is not divisible by $p - m$ or $h(-m)$ is not divisible by $p + m$.

EXAMPLE 1.

$$g(x) = 2x^5 + 7x^4 + 3x^3 - 11x^2 - 16x - 12.$$

Here $h(y) = y^5 + 7y^4 + 6y^3 - 44y^2 - 128y - 192$; the divisors of the number 192 are $\pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 8, \pm 12, \pm 16, \pm 24, \pm 32, \pm 48, \pm 64, \pm 96, \pm 192$; $h(1) = -350$, $h(-1) = -108$; the divisors p of the number 192 for which $h(1)$ is divisible by $p - 1$ and $h(-1)$ is divisible by $p + 1$ are $2, 3, -4, 8$. With the help of Horner's scheme we verify that $h(2) \neq 0$, $h(3) = 0$, $h(-4) = 0$, $h(8) \neq 0$, where -4 is a double root of $h(y)$. The rational roots of $g(x)$ are $x_1 = 3/2$, $x_{2,3} = -2$.

Instead of going at once from the polynomial (3.20) to the polynomial (3.21), we can first, by testing the divisors of the number a_n, find all the integral roots $x_1, \ldots, x_k$ of the polynomial $g(x)$ (every root is written as many times as its multiplicity), and then, in the way given above, we find all the rational roots of the quotient after the division of $g(x)$ by $(x - x_1) \ldots (x - x_k)$.

EXAMPLE 2.

$$g(x) = 3x^6 - 5x^4 - 10x^3 - 8x^2 + x - 2.$$

The number $x = 2$ is a root of $g(x)$ of multiplicity 1, $g(x)$ has no other integral roots. The quotient after the division of $g(x)$ by $x - 2$ is $g_1(x) = 3x^5 + 6x^4 + 7x^3 + 4x^2 + 1$. The multiplication of $g_1(x)$ by 3^4 and the substitution

$y = 3x$ gives $h_1(y) = y^5 + 6y^4 + 21y^3 + 36y^2 + 81$, $h_1(1) = 145$, $h_1(-1) = 101$. Not for any divisor p of the number 81 is $h_1(1)$ divisible by $p - 1$ (nor is $h_1(-1)$ divisible by $p + 1$), i.e. $h_1(y)$ has no integral roots. Hence, $g_1(x)$ has no rational roots. The unique rational root of the polynomial $g(x)$ is $x_1 = 2$.

If for the polynomial (3.20) the numbers a_0, a_n and at least one of the numbers $g(1)$, $g(-1)$ are odd or if the numbers a_0, a_n and both the numbers $g(1)$, $g(-1)$ are not divisible by 3, then the polynomial $g(x)$ does not have any rational (including integral) roots ([11], Ed. 4, p. 352).

EXAMPLE 3. $g(x) = 2x^8 - x^7 + 8x^4 + x^2 - 5$.

Here the numbers $a_0 = 2$, $a_8 = -5$, $g(1) = 5$, $g(-1) = 7$ are not divisible by 3, $g(x)$ has no rational roots.

6. Decomposition of a polynomial into factors of the first and second degree.

If the polynomial $f(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1}x + a_n$ has the roots $\alpha_1, \alpha_2, \ldots, \alpha_s$ of multiplicities $k_1, \ldots, k_s$ respectively $(k_1 + \ldots + k_s = n)$, then it can be decomposed in the following way into the factors:

$$f(x) = a_0(x - \alpha_1)^{k_1}(x - \alpha_2)^{k_2} \ldots (x - \alpha_s)^{k_s} \tag{3.22}$$

(*the decomposition of $f(x)$ into linear factors*). The decomposition of every polynomial into linear factors is unique to within the order of the factors.

If $f(x)$ is a polynomial with real coefficients and α is a complex root of this polynomial, then the number $\bar{\alpha}$ conjugate to α ([11], p. 121) is also a root of $f(x)$, and the multiplicities of the roots α and $\bar{\alpha}$ are the same. If in the decomposition (3.22), given for a polynomial with real coefficients, we multiply together pairs of brackets corresponding to complex conjugate roots of this polynomial, then we obtain a decomposition of $f(x)$ in factors which also have real coefficients:

$$f(x) = a_0(x - \alpha_1)^{k_1} \ldots (x - \alpha_l)^{k_l} \times (x^2 + p_1 x + q_1)^{k_1'} \ldots (x^2 + p_t x + q_t)^{k_t'}.$$

The quadratic trinomials occurring in this expansion do not have real roots and cannot be decomposed into linear factors with real coefficients. Such a decomposition of the polynomial $f(x)$ is unique to within the order of the factors.

EXAMPLE.

$$f(x) = (x - 1)(x - i)(x + i)[x - (1 + i)]^2[x - (1 - i)]^2 = (x - 1)(x^2 + 1)(x^2 - 2x + 2)^2.$$

7. Irreducible polynomials.

The polynomial $f(x)$ of degree $n \geqslant 1$ with coefficients from the field K (p. 152) is called *irreducible* over this field if it cannot be decomposed as the product of polynomials of degree less than n with coefficients from the same field; in the opposite case the polynomial $f(x)$ is called *reducible* over the field K. All polynomials of the first degree are irreducible over any field. If K is the field of all complex numbers, then polynomials of degree above the first, irreducible over this field, do not exist (see the preceding section). Over the field of real numbers, besides the polynomials of the first degree, the irreducible polynomials are also polynomials of the second degree which do not have real roots, and all polynomials of degree above the second are reducible. If K is the field of rational numbers, then for every n there exists a polynomial of degree n which is irreducible over this field.

EXAMPLE 1.

$$f(x) = x^2 + 1.$$

This polynomial is irreducible over the field of real numbers, but is reducible over the field of complex numbers: $f(x) = (x + i)(x - i)$.

EXAMPLE 2.

$$f(x) = x^2 - 2.$$

This polynomial is reducible over the field of complex numbers, and also over the field of real numbers: $f(x) = (x - \sqrt{2})(x + \sqrt{2})$. The given polynomial is irreducible over the field of rational numbers.

A polynomial of above the first degree, irreducible over some field, cannot have roots in this field (the converse is not true).

If $f(x)$ is an arbitrary, and $p(x)$ an irreducible polynomial, both with coefficients in the field K, then either $f(x)$ divides $p(x)$ without remainder, or else these polynomials are mutually prime (section 1, subsection 6). If the product of several polynomials is divisible by an irreducible polynomial $p(x)$, then at least one of the factors is divisible by $p(x)$.

Every polynomial $f(x)$ with coefficients in the field K, having degree $n \geqslant 1$, can be decomposed into the product of polynomials irreducible over this field, and if the polynomial $f(x)$ is decomposed in two ways as the product of irreducible factors $f(x) = p_1(x)\, p_2(x) \ldots p_s(x) = q_1(x)\, q_2(x) \ldots q_t(x)$, then $s = t$ and, in a corresponding numbering, the equations $q_i(x) = c_i\, p_i(x)$ hold, $i = 1, 2, \ldots, s$, where c_i is an element of the field K distinct from zero.

If in the decomposition of the polynomial $f(x)$ into irreducible factors we take the leading coefficient outside the brackets of each of these factors, then we obtain a decomposition $f(x) = a_0\, p_1(x)\, p_2(x) \ldots p_s(x)$, where all the $p_i(x)$, $i = 1, 2, \ldots, s$, are irreducible factors with leading coefficient equal to 1. For every polynomial such a decomposition is completely defined to within the numbering of the factors. If the irreducible polynomial $p(x)$ occurs in the given decomposition of the polynomial $f(x)$ k times, then $p(x)$ is said to be a *k-tuple factor* of $f(x)$.

If we are given decompositions of the polynomials $f(x)$ and $g(x)$ into irreducible factors, then the greatest common divisor $d(x)$ of these polynomials is equal to the product of the factors which occur in both decompositions, where every factor is taken with the degree equal to the smaller of its multiplicities in the two given polynomials.

EXAMPLE 3.

$$f(x) = 2(x + 1)(x^2 + 1)(x^2 - x + 1)^3,$$
$$g(x) = 5(x^2 + 1)^4(x^2 - x + 1)^2(x^2 + 3x + 4)^3$$

are the decompositions of the polynomials into irreducible factors over the field of real numbers. $d(x) = (x^2 + 1) \times (x^2 - x + 1)^2$.

8. Decomposition of polynomials into irreducible factors over the field of rational numbers.

The polynomial $f(x)$ with rational coefficients is irreducible over the field of rational numbers (subsection 7) if and only if the polynomial with integral coefficients, obtained by multiplying $f(x)$ by the least common multiple of the denominators of its coefficients, is irreducible over this field. A polynomial with integral coefficients is irreducible over the field of rational numbers if and only if it cannot be decomposed into factors also with integral coefficients, each of which has degree lower than the polynomial itself.

Let

$$(gx) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n \qquad (3.23)$$

be a polynomial with integral coefficients.

Eisenstein's criterion. If we can find a prime number p such that in the polynomial (3.23) a_0 is not divisible by p, all the remaining coefficients are divisible by p, but a_n, divisible by p, is not divisible by p^2, then this polynomial is irreducible over the field of rational numbers.

EXAMPLE 1.

$$g(x) = 3x^7 - 4x^6 + 2x^5 - 6x^3 - 8x - 2.$$

Here if we use the number $p = 2$, the polynomial $g(x)$ is irreducible over the field of rational numbers.

EXAMPLE 2.

$$g(x) = x^4 + x^3 + x^2 + x + 1.$$

It is not possible to apply Eisenstein's criterion to this polynomial immediately, but we can make the substitution $x = y + 1$, as a result of which we obtain the polynomial $h(y) = y^4 + 5y^3 + 10y^2 + 10y + 5$, which is irreducible by virtue of Eisenstein's criterion ($p = 5$). Hence, the polynomial $g(x)$ is also irreducible over the field of rational numbers.

Polya's criterion ([31], pp. 111, 114). For the polynomial (3.23) let A be the greatest of the numbers $|a_1|, \ldots, |a_n|$ and k an integer with the property $k \geqslant A/|a_0| + 3/2$. Then if $g(k-1) \neq 0$ and $g(k)$ is a prime number, the polynomial $g(x)$ is irreducible over the field of rational numbers.

EXAMPLE 3. $g(x) = x^3 - x^2 + x + 1$.

Here $A = 1$; $k = 3$ is the smallest integer with the property $k \geqslant A/|a_0| + 3/2$; $g(3) = 22$ is a composite number, but $g(4) = 53$ is a prime number. The polynomial $g(x)$ is irreducible over the field of rational numbers.

Cohen's criterion ([31], p. 118). If the coefficients of the polynomial (3.23) satisfy the condition $0 \leqslant a_i \leqslant 9$, $i = 0, 1, \ldots, n$, and $N = a_0 10^n + a_1 10^{n-1} + \ldots + a_{n-1} 10 + a_n$ is a prime number, then this polynomial is irreducible over the field of rational numbers.

EXAMPLE 4. $g(x) = 2x^4 + x^3 + 5x^2 + 6x + 3$; $N = 21,563$ is a prime number; the polynomial $g(x)$ is irreducible.

We will introduce one further criterion of irreducibility ([24], Problem 681). If the polynomial (3.23) takes the values ± 1 for more than $2l$ integral values of x, where $n = 2l$ or $2l + 1$, then it is irreducible over the field of rational numbers.

EXAMPLE 5.

$$g(x) = x^3 - 3x^2 + 3.$$

Here $g(-1) = -1$, $g(1) = 1$, $g(2) = -1$. The polynomial $g(x)$ is irreducible.

For other sufficiency conditions for a polynomial with rational coefficients to be irreducible, see [5], Chapter IV, section 24, and [28], Chapter I, section 4.

For polynomials of the second, third and fourth degrees with rational coefficients we can give not only sufficient, but also necessary conditions for irreducibility over the field of rational numbers ([17], pp. 174–82). A polynomial of the second or third degree with rational coefficients is reducible over the field of rational numbers if and only if it has at least one rational root. A polynomial of the fourth degree

$$f(x) = x^4 + b_1 x^3 + b_2 x^2 + b_3 x + b_4$$

with rational coefficients is reducible over the field of rational numbers if and only if either it has a rational root or else the polynomial of the third degree connected with it

$$k(t) = 8t^3 - 4b_2 t^2 + (2b_4 b_3 - 8b_4)t - (b_1^2 b_4 - 4b_2 b_4 + b_3^2)\ \dagger$$

has a rational root t_0 such that

$$\lambda = \sqrt{\left(2t_0 + \frac{b_1^2}{4} - b_2\right)} \quad \text{and} \quad \mu = \sqrt{(t_0^2 - b_4)}$$

are rational numbers. In the latter case

$$f(x) = \left[x^2 + \left(\frac{b_1}{2} + \lambda\right)x + (t_0 + \mu)\right] \times \left[x^2 + \left(\frac{b_1}{2} - \lambda\right)x + (t_0 - \mu)\right],$$

where the values of the roots λ and μ must be taken with the same sign if $b_1 t_0 - b_3 > 0$, and with opposite signs if $b_1 t_0 - b_3 < 0$.

EXAMPLE 6. $f(x) = x^4 - 4x^3 + 5x^2 - 2$.

Here $x = 1$ is a root of $f(x)$, $f(x) = (x - 1)(x^3 - 3x^2 + 2x + 2)$. The polynomial $x^3 - 3x^2 + 2x + 2$ has no rational roots (subsection 5) and is irreducible over the field of rational numbers.

EXAMPLE 7.

$$f(x) = x^4 - 2x^3 - 7x^2 - 10x - 2.$$

There are no rational roots. The polynomial $k(t) = 8t^3 + 28t^2 + 56t - 36$ has a rational root $t_0 = 1/2$ and $\lambda = \sqrt{(1 + 1 + 7)} = \pm 3$, $\mu = \sqrt{(1/4 + 2)} = \pm 3/2$ are rational numbers; $b_1 t_0 - b_3 = -1 + 10 > 0$, $f(x) = (x^2 + 2x + 2)(x^2 - 4x - 1)$.

EXAMPLE 8.

$$f(x) = 2x^4 - 3x^3 - x^2 - 3x + 1.$$

---

† If we form the polynomial $k(t)$ for the left-hand side of equation (3.16) and make the substitution $t = (z + p)/2$ in it, then we obtain the left-hand side of the cubic resolvent (3.17) of the equation (3.16).

In the way explained in subsection 5, we determine that $f(x)$ has no rational roots. We divide all the coefficients of $f(x)$ by the leading coefficient $a_0 = 2$ and form the polynomial

$$k(t) = 8t^3 - 4\left(-\frac{1}{2}\right)t^2 + \left[2\left(-\frac{3}{2}\right)\left(-\frac{3}{2}\right) - 8 \times \frac{1}{2}\right]t$$
$$- \left[\frac{9}{4}\frac{1}{2} - 4\left(-\frac{1}{2}\right)\frac{1}{2} + \frac{9}{4}\right] = 8t^3 + 2t^2 + \frac{1}{2}t - \frac{35}{8}.$$

Multiplying $k(t)$ by 8 and making the substitution $v = 4t$, we obtain the polynomial $v^3 + v^2 + v - 35$, which does not have rational roots. Hence $k(t)$ also does not have rational roots. The polynomial $f(x)$ is irreducible over the field of rational numbers.

If it is impossible to apply any of the criteria of irreducibility given above to the polynomial (3.23), then the reducibility or not of this polynomial over the field of rational numbers can be established in the following way ([27], pp. 66–68, or [13], pp. 173–6).

We give the variable x m integral values $x_1, x_2, \ldots, x_m$, where $m = n/2 + 1$ for even n and $m = (n+1)/2$ for odd n. Then we form all possible sets of integers of the form $c_1, c_2, \ldots, \ldots, c_m$, where $c_i (i = 1, 2, \ldots, m)$ is a divisor of the number $g(x_i)$ (altogether we obtain $S = 2^m s_1 \ldots s_m$ sets, where s_i is the number of all the positive divisors of $g(x_i)$). For every such set we construct the polynomial $h_j(x)$ $(j = 1_i, 2, \ldots, S)$ of degree $\leqslant m - 1$ with the properties $h_j(x_i) = c_i$, where $i = 1, 2, \ldots, m$, and the numbers c_i are taken from the jth set (section 2). After this, for those of the polynomials $h_j(x)$ all of whose coefficients are integers and whose degrees are greater than 0, we can find by direct division whether $g(x)$ is divisible by $h_j(x)$. If $g(x)$ is not divisible by any of these polynomials, it is irreducible over the field of rational numbers. In the opposite case, we obtain a decomposition of the polynomial $g(x)$ into factors of lower degree with rational coefficients, which are easy to decompose into factors with integral coefficients, with which we can deal in the same way as with the polynomial $g(x)$ before.

In applying the process described above to the polynomial $g(x)$ the following simplifications are possible:

if two sets of numbers $c_1, c_2, \ldots, c_m$ and $c_1', c_2', \ldots, c_m'$ differ from each other only in sign, then the polynomial $h_j(x)$ is constructed for only one of these sets;

we can take several more integral values of the variable $x_{m+1}, x_{m+2}, \ldots, x_{m+l}$; if for at least one of them $g(x_{m+k})$ is not divisible by $h_j(x_{m+k})$ then it is not necessary to verify whether $g(x)$ is divisible by $h_j(x)$.

EXAMPLE 9. $g(x) = x^5 - 2x^4 - 3x^3 + 6x^2 + x - 1$.

Here $g(0) = -1, g(1) = 2, g(-1) = 4, g(2) = 1$. We choose $g(0)$, $g(1)$, $g(2)$ as having the least divisors, and we form the sets of divisors of the numbers chosen:

$$\begin{array}{l|r|r|r|r||r|r|r|r||} c_1 = 1 & 1 & 1 & -1 & 1 & 1 & 2 & -1 \\ c_2 = 1 & 1 & -1 & 1 & 2 & 2 & -2 & 2 \\ c_3 = 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 \end{array} \dagger$$

By Newton's formula (section 2) we form for these sets the polynomials $h_1(x) \equiv 1$, $h_2(x) = -x^2 + x + 1$, $h_3(x) = 2x^2 - 4x + 1$, $h_4(x) = -x^2 + 3x - 1$. $h_5(x) = -x^2 + 2x + 1$, $h_6(x) = -2x^2 + 3x + 1$, $h_7(x) = 3x^2 - 6x + 1$, $h_8(x) = -2x^2 + 5x - 1$, where $h_j(x)$ $(j = 1, 2, \ldots, 8)$ are polynomials which when $x = 0, 1, 2$ take the values standing in the jth column of the table above. The numbers $h_3(-1) = 7$, $h_4(-1) = -5$, $h_7(-1) = 10$ and $h_8(-1) = -8$ are not divisors of the number $g(-1) = 4$, therefore the polynomials $h_3(x)$, $h_4(x)$, $h_7(x)$ and $h_8(x)$ will not be considered. We take the value $g(-2) = -19$. It is not divisible by $h_2(-2) = -5$, $h_5(-2) = -7$, and $h_6(-2) = -13$. The polynomial $g(x)$ is irreducible over the field of rational numbers.

EXAMPLE 10.

$$g(x) = x^4 + x^3 - 6x^2 - 3x + 9.$$

$$g(-2) = -1,\ g(1) = 2,\ g(2) = 3.$$

We form the sets of divisors of these numbers:

$$\begin{array}{l|r|r|r|r||r|r|r|r||r|r|r|r||} c_1 = 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 \\ c_2 = 1 & 1 & -1 & 1 & 2 & 2 & -2 & 2 & 1 & 1 & -1 & 1 \\ c_3 = 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & 3 & 3 & 3 & -3 \end{array}$$

$$\begin{array}{||r|r|r|r||} 1 & -1 & 1 & 1 \\ 2 & 2 & -2 & 2 \\ 3 & 3 & 3 & -3 \end{array} \dagger$$

We construct the polynomials corresponding to them (section 2, Newton's formula):

$h_1(x) \equiv 1$; $h_2(x)$, $h_3(x)$, $h_4(x)$, $h_5(x)$, $h_6(x)$ are polynomials with fractional coefficients; $h_7(x) = x^3 - 3$ is a divisor of $g(x)$, $g(x) = (x^2 - 3)(x^2 + x - 3)$; the polynomials $x^2 - 3$ and $x^2 + x - 3$ are irreducible over the field of rational numbers (because they have no rational roots).

For another way of decomposing a polynomial with rational coefficients into factors of lower degree with rational coefficients, see [27], p. 68.

9. Rational functions.

A *rational function* (or simply a *rational fraction*) is an expression of the form $f(x)/g(x)$, where $f(x)$ and $g(x)$ are poly-

† We do not take the sets which differ from those written here only in sign.

nomials over some field K (section 1), and $g(x) \neq 0$. We can perform algebraic operations on rational functions in the same way as on rational numbers. The rational fraction $f_1(x)/g_1(x)$ is called *irreducible* if the polynomials $f_1(x)$ and $g_1(x)$ are mutually prime (section 1, subsection 6). Every rational fraction is equal to an irreducible fraction, obtained from it by dividing its numerator $f(x)$ and denominator $g(x)$ by a greatest common divisor of $f(x)$ and $g(x)$.

EXAMPLE 1.

$$\frac{x^2 - x}{x^3 + 2x^2 - x} = \frac{x - 1}{x^2 + 2x - 1}.$$

A rational fraction is called *proper* if the degree of its numerator is less than the degree of its denominator. Every rational fraction which is not a polynomial (i.e. such that in its irreducible form the denominator does not have zero degree), can be represented in a unique way in the form of a sum of some polynomial and some proper fraction. In order to obtain such a representation of the fraction $f(x)/g(x)$, it is necessary to divide the polynomial $f(x)$ by $g(x)$ (section 1, subsection 2); if in doing this we obtain the quotient $q(x)$ and the remainder $r(x)$, then we will have

$$\frac{f(x)}{g(x)} = q(x) + \frac{r(x)}{g(x)},$$

where $r(x)/g(x)$ is a proper rational fraction.

EXAMPLE 2.

$$\frac{f(x)}{g(x)} = \frac{3x^4 + 2x^2 - 4x + 6}{x^2 - 2x + 3}.$$

Here

$$3x^4 + 2x^2 - 4x + 6 = (x^2 - 2x + 3)(3x^2 + 6x + 5) - 12x - 9,$$

$$\frac{f(x)}{g(x)} = 3x^2 + 6x + 5 + \frac{-12x - 9}{x^2 - 2x + 3}.$$

A proper rational fraction $f(x)/g(x)$ is called a partial fraction if $g(x)$ is a power of some polynomial $p(x)$, irreducible (subsection 7) over the given field K ($g(x) = p^k(x)$, where $k \geqslant 1$), and the degree of the polynomial $f(x)$ is less than the degree of $p(x)$. If K is the field of all complex numbers, then a partial fraction is a fraction of the form $a/(x-b)^l$ where a and b are complex numbers and $l \geqslant 1$; if K is the field of all real numbers, then the partial fractions are the fractions $a/(x-b)^l$ (a, b are real numbers, $l \geqslant 1$) and $(a_1 x + a_2)/(x^2 + b_1 x + b_2)^m$, where $x^2 + b_1 x + b_2$ is a polynomial with real

coefficients which does not have real roots, a_1, a_2 are real numbers, $m \geqslant 1$.

The fundamental theorem. Every proper rational fraction can be decomposed into the sum of partial fractions, and this decomposition is unique.

In order to obtain the decomposition of the fraction $f(x)/g(x)$ into partial fractions over the given number field K, we must decompose the polynomial $g(x)$ into factors irreducible over K and, if we obtain $g(x) = p_1^{s_1}(x) \dots p_t^{s_t}(x)$ ($p_1(x), \dots, p_t(x)$ are irreducible polynomials), we write

$$\frac{f(x)}{g(x)} = \frac{u_1^{(1)}(x)}{p_1^{s_1}(x)} + \frac{u_2^{(1)}(x)}{p_1^{s_1-1}(x)} + \dots + \frac{u_{s_1}^{(1)}(x)}{p_1(x)} + \dots$$
$$\dots + \frac{u_1^{(t)}(x)}{p_t^{s_t}(x)} + \frac{u_2^{(t)}(x)}{p_t^{s_t-1}(x)} + \dots + \frac{u_{s_t}^{(t)}(x)}{p_t(x)}, \qquad (3.24)$$

where $u_j^{(i)}(x)$ ($i = 1, \dots, t; j = 1, \dots, s_i$ for fixed i) is a polynomial of lower degree than $p_i(x)$, taken with non-negative coefficients. Then we must reduce all the fractions on the right-hand side of equation (3.24) to a common denominator $g(x)$ and equate the sum obtained in this way in the numerator to the polynomial $f(x)$. Equating coefficients of similar powers of x on the left- and right-hand sides of the equation obtained (or equating the variable x on the left- and right-hand sides of the equation to numerical values), we obtain a system of linear equations whose solution (Chapter I, section 2, subsections 1–3) gives us the coefficients of the polynomials $u_j^{(i)}(x)$.

EXAMPLE 3. The rational fraction

$$\frac{f(x)}{g(x)} = \frac{2x^4 - 2x^3 + 6x^2 + 2x + 7}{x^5 + 2x^4 + 2x^3 + 4x^2 + x + 2}$$

is to be decomposed into partial fractions over the field of real numbers.

Here $g(x) = (x + 2)(x^2 + 1)^2$, where the polynomials $x + 2$ and $x^2 + 1$ are irreducible. We write

$$\frac{f(x)}{g(x)} = \frac{a}{x + 2} + \frac{a_1 x + a_2}{(x^2 + 1)^2} + \frac{c_1 x + c_2}{x^2 + 1}.$$

Reducing all the fractions to the common denominator $g(x)$, we obtain the equation

$$f(x) = a(x^2 + 1)^2 + (a_1 x + a_2)(x + 2) + (c_1 x + c_2)(x + 2)(x^2 + 1),$$

or

$$2x^4 - 2x^3 + 6x^2 + 2x + 7 = (a + c_1)x^4 + (2c_1 + c_2)x^3$$
$$+ (2a + a_1 + c_1 + 2c_2)x^2 + (2a_1 + a_2 + 2c_1 + c_2)x + (a + 2a_2 + 2c_2).$$

Equating coefficients of similar powers of x, we obtain the system of equations

$$\begin{cases} a + c_1 = 2, \\ 2c_1 + c_2 = -2, \\ 2a + a_1 + c_1 + 2c_2 = 6, \\ 2a_1 + a_2 + 2c_1 + c_2 = 2, \\ a + 2a_2 + 2c_2 = 7. \end{cases}$$

Solving this system, we find: $a = 3, a_1 = 1, a_2 = 2, c_1 = -1, c_2 = 0$. Thus,

$$\frac{f(x)}{g(x)} = \frac{3}{x+2} + \frac{x+2}{(x^2+1)^2} - \frac{x}{x^2+1}.$$

EXAMPLE 4. The same rational fraction is to be decomposed into partial fractions over the field of complex numbers.

Here $g(x) = (x+2)(x+i)^2(x-i)^2$. Corresponding to this we write

$$\frac{f(x)}{g(x)} = \frac{a}{x+2} + \frac{b}{(x+i)^2} + \frac{c}{x+i} + \frac{d}{(x-i)^2} + \frac{e}{x-i}.$$

We reduce all the fractions to the common denominator $g(x)$ and obtain the equation

$$2x^4 - 2x^3 + 6x^2 + 2x + 7 = a(x+i)^2(x-i)^2 + b(x+2)(x-i)^2 + c(x+2)(x+i)(x-i)^2 + d(x+2)(x+i)^2 + e(x+2)(x+i)^2(x-i).$$

Taking the value $x = -2$, we obtain $75 = a \cdot 25$, i.e. $a = 3$. We take $x = -i$. This gives $3 - 4i = b(2-i)(-4)$, whence $b = -1/2 + 1/4\,i$. When $x = i$ we obtain $3 + 4i = d(2+i)(-4)$, i.e. $d = -1/2 - 1/4\,i$. Now we let $x = 0$ and then $x = -1$ and, using the values of the coefficients a, b, d already found, we obtain

$$\begin{cases} 7 = 3 + \left(1 - \frac{1}{2}\,i\right) - 2ci + \left(1 + \frac{1}{2}\,i\right) + 2ei, \\ 15 = 12 + \left(-\frac{1}{2} - i\right) + c(-2-2i) + \left(-\frac{1}{2} + i\right) + e(-2+2i), \end{cases}$$

or

$$\begin{cases} -2ci + 2ei = 2, \\ c(-2-2i) + e(-2+2i) = 4, \end{cases}$$

whence

$$c = \frac{-1+i}{2}, \quad e = \frac{-1-i}{2}.$$

Thus

$$\frac{f(x)}{g(x)} = \frac{3}{x+2} - \frac{2-i}{4(x+i)^2} - \frac{1-i}{2(x+i)} - \frac{2+i}{4(x-i)^2} - \frac{1+i}{2(x-i)}.$$

EXAMPLE 5. To decompose the fraction $f(x)/g(x) = (x^2+2)/(x-1)^3$ into partial fractions.

We write

$$\frac{f(x)}{g(x)} = \frac{a}{(x-1)^3} + \frac{b}{(x-1)^2} + \frac{c}{x-1}.$$

We reduce all the fractions to the common denominator and obtain $f(x) = a + b(x-1) + c(x-1)^2$. Setting in turn $x=1$, $x=0$, $x=2$, we find $a=3$, $b=2$, $c=1$. Thus:

$$\frac{x^2+2}{(x-1)^3} = \frac{3}{(x-1)^3} + \frac{2}{(x-1)^2} + \frac{1}{x-1}.$$

For more on the decomposition of fractions into simple fractions, see section 1, subsection 4 and section 2 (Lagrange's formula).

4. Bounds on the Roots of a Polynomial. Theorems on the Number of Roots

1. Bounds on the real roots of a polynomial with real coefficients.

If

$$f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n \quad (a_0 \neq 0) \quad (3.25)$$

is a polynomial with real coefficients, then all its real roots (if they exist) are contained in the interval $(-N, N)$, where $N = 1 + \frac{A}{|a_0|}$, A is the largest of the numbers $|a_1|, \ldots, |a_n|$ (cf. section 4, subsection 3).

EXAMPLE 1.

$$f(x) = 5x^6 - 11x^5 + 38x^4 + x^3 - 100x^2 - 3.$$

Here $A = 100$, $N = 1 + 100/5 = 21$.

We can give even more precise bounds for the roots of the polynomial (3.25). The number K is called an *upper bound of the real roots* of the polynomial (3.25), if this polynomial does not have any real roots greater than or equal to K; similarly, we define *upper* and *lower bounds of the positive* or *negative roots* of the polynomial. If K_1 is an upper bound of the positive roots of $f(x)$, K_2 an upper bound of the positive roots of $f(-x)$, $K_3 > 0$ an upper bound of the positive roots of the polynomial $x^n f(1/x)$† and $K_4 > 0$ is an upper bound of the positive roots of the polynomial $x^n f(-1/x)$, then all

† This polynomial can be obtained by writing all the coefficients of the polynomial $f(x)$ in the reverse order:

$$x^n f\left(\frac{1}{x}\right) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0.$$

the real roots of the polynomial $f(x)$ distinct from zero, if they exist, lie inside the intervals $(-K_2, -1/K_4)$ and $(1/K_3, K_1)$.

In order to determine an upper bound of the positive roots of a polynomial we can apply one of the following methods.

(1) If in the polynomial (3.25) $a_0 > 0$, $a_1, a_2, \ldots, a_{l-1} \geqslant 0$ and $a_l < 0$, then an upper bound of the positive roots of this polynomial is found by means of the formula

$$K = 1 + \sqrt[l]{\frac{B}{a_0}},$$

where B is the greatest of the absolute values of the negative coefficients of $f(x)$ (*Maclaurin's method*).

EXAMPLE 2.

$$f(x) = 9x^4 - 9x^2 - 36x + 1.$$

Here $l = 2$, $B = 36$, $K = K_1 = 1 + \sqrt{(36/9)} = 3$. We consider the polynomial $f(-x) = 9x^4 - 9x^2 + 36x + 1$. For it an upper bound of the positive roots is the number $K_2 = 1 + \sqrt{\frac{9}{9}} = 2$. We take $x^4 f(1/x) = x^4 - 36x^3 - 9x^2 + 9$. Here $l = 1$, $K_3 = 1 + 36 = 37$. Finally, for $x^4 f(-1/x) = x^4 + 36x^3 - 9x^2 + 9$ we have $K_4 = 1 + \sqrt{9} = 4$. Hence, if the polynomial $f(x)$ has real roots, then they must necessarily lie in the intervals $(-2, -1/4)$ and $(1/37, 3)$ (by the method described at the beginning of this section, we would obtain the interval $(-5, 5)$).

(2) If when $x = c$ the polynomial (3.25) and its derivatives $f'(x), f''(x), \ldots, f^{(n)}(x)$ take positive values, then c is an upper bound of the positive roots of $f(x)$ (*Newton's method*). In applying Newton's method, the values $f(c), f'(c), \ldots, f^{(n)}(c)$ are most easily found by Horner's scheme (section 1, subsections 3–4).

EXAMPLE 3.

$$f(x) = 3x^4 - 18x^3 + 14x^2 + 36x + 25.$$

By Maclaurin's method we find $K = 1 + 18/3 = 7$. We apply Newton's method. We test the number $c = 3$;

	3	−18	14	36	25
3	3	−9	−13	−3	16
	3	0	−13	−42	

We get that $f(3) = 16 > 0$, but $f'(3) = -42 < 0$; $c = 3$ is not suitable. We test $c = 4$:

	3	−18	14	36	25
4	3	−6	−10	−4	9
	3	6	14	52	

Because we have obtained a row consisting entirely of positive numbers, we need not reckon any further (everywhere else we will obtain positive numbers). $f(4), f'(4), f''(4), f'''(4), f^{IV}(4) > 0$, $c = 4$ is an upper bound of the positive roots of $f(x)$.

To find the number c for which $f(c) > 0$, $f'(c) > 0, \ldots,$ $f^n(c) > 0$, we can also proceed as follows. Take a number c_1 for which $f^{n-1}(c_1) > 0$, then a number $c_2 \geqslant c_1$ for which $f^{n-2}(c_2) > 0$, then a number $c_3 \geqslant c_2$ for which $f^{n-3}(c_3) > 0$, and so on until $c_n = c$.

EXAMPLE 4.

$$f(x) = x^3 - 6x^2 + 7x + 5.$$

Here $f'(x) = 3x^2 - 12x + 7$, $f''(x) = 6x - 12$ and we can take $c_1 = 3$, $c_2 = 4$, $c_3 = 4$. $c = c_3 = 4$ is an upper bound of the positive roots of $f(x)$.

(3) Presupposing $a_0 > 0$, we represent the polynomial (3.25) (without changing the position of its terms) in the form $f(x) = f_1(x) + \ldots + f_s(x)$, where every polynomial $f_i(x)$ $(i = 1, \ldots, s)$ has positive leading coefficient and in the series of coefficients has only one change of sign† ($f_s(x)$ in general cannot have a change of sign in its series of coefficients). If, when $c > 0$, $f_1(c) > 0, \ldots, f_s(c) > 0$, then c is an upper bound of the positive roots of $f(x)$ ([31], p. 170).

EXAMPLE 5.

$$f(x) = x^7 + 2x^6 - 3x^5 - 3x^4 + x^3 - 2x^2 + x + 2.$$

We set

$$f_1(x) = x^7 + 2x^6 - 3x^5 - 3x^4, \quad f_2(x) = x^3 - 2x^2, \quad f_3(x) = x + 2.$$

Then $f_1(3) > 0$, $f_2(3) > 0$, $f_3(3) > 0$ and $c = 3$ is an upper bound of the positive roots of $f(x)$.

If all the coefficients of the polynomial $f(x)$ are positive, it does not in general have positive roots and zero or any positive number can be regarded as an upper bound for its positive roots.

If, in the polynomial (3.25), $a_0 < 0$, then it is necessary to consider the polynomial $-f(x)$ which has the same roots as $f(x)$.

EXAMPLE 6.

$$f(x) = -2x^4 - 3x^3 + 5x^2 + 6.$$

Here $-f(x) = 2x^4 + 3x^3 - 5x^2 - 6$. By Maclaurin's method for the polynomial $-f(x)$ we obtain $K = 1 + \sqrt{(6/2)} = 1 + \sqrt{3} \approx 2{\cdot}74$ (we take the approx-

† See footnote on p. 193.

imate value of $\sqrt{3}$ corrected upwards). Proceeding by Newton's method, we test $c = 2$:

	2	3	−5	0	−6
2	2	7	9	18	30

We have obtained a row of positive numbers, so that $c = 2$ is an upper bound of the positive roots of the polynomial $-f(x)$, i.e. of the polynomial $f(x)$.

2. Theorems on the number of real roots.

If every root of a polynomial is considered as many times as its multiplicity (section 1, subsection 10), then the number of all the real roots of the polynomial $f(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n$ with real coefficients always has the same parity as the degree n of the polynomial. In particular, if n is odd, then $f(x)$ necessarily has at least one real root. The number of positive roots is even or odd depending on whether a_0 and the last non-zero coefficient of $f(x)$ have the same or opposite signs.

If, when $x = b$ and $x = c$, $f(x)$ takes values with opposite signs, then there is at least one real root of the polynomial $f(x)$, in the segment $[b, c]$, and the total number of roots of $f(x)$ in the segment $[b, c]$ is odd (if again every root is counted as many times as its multiplicity). If the values of $f(b)$ and $f(c)$ have the same sign, then in general either $f(x)$ does not have any roots in the segment $[b, c]$, or else has an even number of them.

The number of all the negative roots of the polynomial $f(x)$ is equal to the number of all the positive roots of the polynomial $f(-x)$.

Descartes' theorem. The number of positive roots of the polynomial $f(x)$ with real coefficients (each counted as many times as its multiplicity) is either equal to the number of changes of sign† in the series of coefficients of $f(x)$, or is an even number less.

EXAMPLE 1.

$$f(x) = x^3 - 2x + 1.$$

The signs of the coefficients are $+\ -\ +$, i.e. we have two changes of sign. The number of positive roots of the polynomial $f(x)$ is either 2 or 0. $f(-x) = -x^3 + 2x + 1$. Here there is one change of sign in the series of coefficients. The number of negative roots of $f(x)$ is equal to 1.

† If in a series of numbers, arranged in a definite order, there are k transitions from one sign to the other (we ignore numbers equal to zero), then we say that this series contains k changes of sign.

Descartes' theorem enables us to determine exactly the number of positive roots of a polynomial, if in the series of its coefficients there are no changes of sign or there is one change of sign. If there is more than one, then sometimes we are able to determine the exact number of positive roots of a polynomial by applying Descartes' theorem and considering in addition the signs of the values of the polynomial at separate points, or else the behaviour of separate terms of the polynomial as x varies.

EXAMPLE 2.

$$f(x) = x^4 - 2x^3 - 2x + 1.$$

In the series of coefficients of $f(x)$ there are two changes of sign, i.e. $f(x)$ has either 2 or 0 positive roots. $f(0) = 1 > 0$, $f(1) = -2 < 0$, i.e. the segment $[0, 1]$ contains at least one root of the polynomial $f(x)$. Hence, the number of positive roots of $f(x)$ is equal to two. $f(-x) = x^4 + 2x^3 + 2x + 1$. The polynomial $f(x)$ has no negative roots.

EXAMPLE 3.

$$f(x) = x^7 + x^2 - x + 1.$$

In the series of coefficients there are two changes of sign. But when $0 < x \leqslant 1$ the only negative term of the polynomial, $-x$, is "covered" by the free term 1, and when $x > 1$ the term x^2 is greater than x. Hence, for any positive value of x the polynomial $f(x)$ takes positive values and the number of positive roots of the polynomial $f(x)$ is equal to 0.

Sometimes the exact number of all the real roots of a polynomial with real coefficients can be determined by applying Descartes' theorem and considering in more detail the graph of this polynomial in the sections included between the bounds of its real roots.

EXAMPLE 4.

$$f(x) = x^5 - 2x^4 - 8x^3 - x^2 - 9x + 1.$$

By Descartes' theorem, we have either 2 or 0 positive roots. $f(0) > 0$, $f(1) < 0$, therefore there are two positive roots. We take $f(-x) = -x^5 - 2x^4 + 8x^3 - x^2 + 9x + 1$. Here there are three changes of sign, i.e. $f(x)$ has either one or three negative roots. By the method of subsection 1 we determine that all the negative roots of the polynomial $f(x)$ lie in the segment $[-3, -2]$. In the same way we determine that -2 is a lower bound of the real roots of the derivative $f'(x)$, i.e. in the segment $[-3, -2] f(x)$ is a monotonic function ([34], p. 53). Hence, $f(x)$ has one negative root. The total number of real roots of the polynomial $f(x)$ is 3.

In some cases the precise number of all the real roots of the polynomial can be determined by applying Descartes'

theorem and calculating the discriminant $D(f)$ of this polynomial (section 1, subsection 12).

EXAMPLE 5.

$$f(x) = x^4 + x^2 - 4x + 3.$$

With the help of Descartes' theorem we find that the number of positive roots of $f(x)$ is 2 or 0, and that there are no negative roots, i.e. either all the roots of the polynomial $f(x)$ are complex, or it has two real and two complex roots. $D(f) = 5744 > 0$, i.e. the number of pairs of complex conjugate roots of $f(x)$ is even (section 1, subsection 12). Hence all the roots of the polynomial $f(x)$ are complex.

A stronger form of Descartes' theorem. If all the roots of the polynomial $f(x)$ are real, then the number of positive roots of $f(x)$ is equal to the number of changes of sign in the series of its coefficients, and the number of negative roots is equal to the number of changes of sign in the series of coefficients of the polynomial $f(-x)$.

EXAMPLE 6.

$$f(x) = \begin{vmatrix} -2-x & 1 & 1 \\ 1 & 1-x & 3 \\ 1 & 1 & 1-x \end{vmatrix} = -x^3 + 14x + 20.$$

Here $f(x)$ is the characteristic polynomial of a symmetric matrix (Chapter II, section 1, subsection 17), therefore all the roots are real. $f(-x) = x^3 - 14x + 20$. The number of positive roots of the polynomial $f(x)$ is equal to 1. The number of negative roots is equal to 2.

If two neighbouring coefficients of the polynomial $f(x)$ are zero, but $a_n \neq 0$, then it is known that not all the roots of $f(x)$ are real ([23], p. 151); in particular, it is impossible to apply the strengthened form of Descartes' theorem.

The number of real roots of the polynomial $f(x)$ of degree n with real coefficients which are greater than a fixed real number c is either equal to the number of changes of sign in the series $f(c), f'(c), \ldots, f^{(n)}(c)$, or is an even number less than this. If all the roots of the polynomial $f(x)$ are real, then the number of roots greater than c is exactly equal to the number of changes of sign in the series $f(c), f'(c), \ldots, f^{(n)}(c)$.

EXAMPLE 7.

$$f(x) = x^4 - 3x^3 + x^2 - 2x + 1.$$

Here $f(1) = -2$, $f'(1) = -5$, $f''(1) = 4$, $f'''(1) = 6$, $f^{IV}(1) = 24$. We have obtained one change of sign, the polynomial $f(x)$ has one real root, greater than 1.

The Budan–Fourier theorem. If the real numbers a and b $(a < b)$ are not roots of the polynomial $f(x)$ of degree n with real coefficients, then the number of real roots of this polynomial lying between a and b, and counted as many times as their multiplicity, is either equal to or an even number less than the difference between the number S_1 of changes of sign in the series of numbers $f(a), f'(a), \ldots, f^{(n)}(a)$ and the number S_2 of changes of sign in the series obtained from the series $f(b), f'(b), \ldots, f^{(n)}(b)$ by replacing the zeros occurring in it by numbers distinct from zero such that if $f^{(k-1)}(b) \neq 0$, $f^{(k)}(b) = \ldots = f^{(k+l-1)}(b) = 0$, $f^{(k+l)}(b) \neq 0$, then the number c_{k+i} replacing $f^{(k+i)}(b)$ $(i = 0, 1, \ldots, l-1)$ has the same sign as $f^{(k+l)}(b)$ if the difference $l - i$ is even, and the opposite sign if this difference is odd.

EXAMPLE 8.

$$f(x) = 3x^8 - 2x^5 + x + 4x^2 - x - 1.$$

To determine the number of roots of $f(x)$ in the segment $[-2, 0]$.

We have: $f(-2) > 0$, $f'(-2) < 0$, $f''(-2) > 0$, $f'''(-2) < 0$, $f^{(4)}(-2) > 0$, $f^{(5)}(-2) < 0$, $f^{(6)}(-2) > 0$, $f^{(7)}(-2) < 0$, $f^{(8)}(-2) > 0$, whence $S_1 = 8$; $f(0) < 0$, $f'(0) < 0$, $f''(0) > 0$, $f'''(0) = 0$, $f^{(4)}(0) > 0$, $f^{(5)}(0) < 0$, $f^{(6)}(0) = 0$, $f^{(7)}(0) = 0$, $f^{(8)}(0) > 0$; we form the series of signs $- - + - + - + - +$; we obtain $S_2 = 7$. $S_1 - S_2 = 1$, one root of the polynomial $f(x)$ lies in the segment $[-2, 0]$.

Sturm's method of determining the number of real roots. The sequence of non-zero polynomials

$$f(x) = f_0(x), \quad f_1(x), \ldots, f_m(x) \tag{3.26}$$

with real coefficients is called a *Sturm series* for the polynomial $f(x)$ if:

(1) when the real variable x increases through a root of the polynomial $f(x)$, the product $f(x) \cdot f_1(x)$ changes sign from $-$ to $+$;

(2) no two neighbouring polynomials $f_k(x)$, $f_{k+1}(x)$ $(k = 0, 1, \ldots, m-1)$ have real roots in common;

(3) if, for some real number α and some index k $(1 \leqslant k \leqslant m-1)$, $f_k(\alpha) = 0$, then the numbers $f_{k-1}(\alpha)$ and $f_{k+1}(\alpha)$ have different signs;

(4) $f_m(x)$ does not have real roots.

The sequence of polynomials (3.26) is called a *Sturm series* for the polynomial $f(x)$ *in the segment* $[a, b]$ if the properties (1), (2), (3) and (4) are satisfied in this segment ([23], p. 141).

Sturm's theorem. Let $f(x)$ be a polynomial with real coefficients, a and b $(a < b)$ real numbers which are not roots of

it, (3.26) a Sturm series for $f(x)$ in the segment [a, b]. Then the number of distinct real roots of the polynomial $f(x)$ lying in the segment $[a, b]$ is equal to the difference between the number of changes of sign† $W(a)$ in the series of values $f(a)$, $f_1(a), \ldots, f_m(a)$ and the number of changes of sign $W(b)$ in the series of values $f(b)$, $f_1(b), \ldots, f_m(b)$.

The construction of a Sturm series. We take the polynomial $f(x) = f_0(x)$, its derivative $f'(x)$, $= f_1(x)$, then the remainder $r_1(x)$ obtained when $f(x)$ is divided by $f'(x)$, with the opposite sign $(-r_1(x) = f_2(x))$, then the remainder $r_2(x)$, obtained when $f'(x)$ is divided by $f_2(x)$, with the opposite sign $(-r_2(x) = f_3(x))$, the remainder after the division of $f_2(x)$ by $f_3(x)$ with the opposite sign $(-r_3(x) = f_4(x))$ and so on up to the last remainder not equal to zero, $r_{m-1}(x)$, $(-r_{m-1}(x) = f_m(x))$. If $f_m(x)$ does not have real roots, then the series of polynomials constructed is a Sturm series for the polynomial $f(x)$; if $f_m(x)$ has real roots, but all of them lie outside the segment $[a, b]$, then this is a Sturm series for the polynomial $f(x)$ in the segment $[a, b]$. If the polynomial $f_m(x)$ has real roots in the segment $[a, b]$, then it is necessary to divide the polynomials $f_0(x), f_1(x), \ldots, f_m(x)$ by $f_m(x)$ and take the quotients obtained. This will be a Sturm series for the polynomial $q_0(x) = f_0(x)/f_m(x)$ ([23], p. 145), which has exactly the same roots as $f(x)$, but of multiplicity 1 (section 1, subsection 10); the difference $W(a) - W(b)$ (a, b are real numbers, which are not roots of $f(x)$), calculated for this series, gives the number of distinct real roots of the polynomial $f(x)$ in the segment $[a, b]$; here in fact division by $f_m(x)$ cannot be carried out, because for any real value $x = c$, which is not a root of the polynomial $f(x)$, the number of changes of sign in the series of values of the quotients obtained coincides with the number of changes of sign in the series of values $f_0(c), f_1(c), \ldots, f_m(c)$.

EXAMPLE 9. To find the number of distinct real roots of the polynomial $f_0(x) = x^4 - 3x^3 - x^2 + 8x - 4$ situated in the segment $[-1, 1]$.

We take the polynomials $f_0(x) = f(x), f_1(x) = f'(x) = 4x^3 - 9x^2 - 2x + 8$, $f_2(x) = -r_1(x) = 35x^2/16 - 45x/8 + 5/2$, $f_3(x) = -r_2(x) = 160x/49 - 320/49$ ($f_2(x)$ is divisible by $f_3(x)$ without remainder). Because in the segment $[-1, 1]$ the polynomial $f_3(x)$ does not have any roots, we have obtained a Sturm series for the polynomial $f(x)$ in the segment $[-1, 1]$. We form a table, in whose first column we write the values which the variable x takes, in the following columns we write the signs obtained for these values of the poly-

† See footnote on p. 193.

nomials in the Sturm series, and in the last column the number of changes of sign

x	$f_0(x)$	$f_1(x)$	$f_2(x)$	$f_3(x)$	W
-1	$-$	$-$	$+$	$-$	2
1	$+$	$+$	$-$	$-$	1

$W(-1) - W(1) = 2 - 1 = 1$. $f(x)$ has one root in the segment $[-1, 1]$.

EXAMPLE 10. To determine the number of distinct roots of the polynomial $f(x)$ in the segment $[-5, 5]$. The series of polynomials constructed above is not a Sturm series in this segment, because $f_3(x)$ has a root $x = 2$ in this segment. Nevertheless, we write out the table:

x	$f_0(x)$	$f_1(x)$	$f_2(x)$	$f_3(x)$	W
-5	$+$	$-$	$+$	$-$	3
5	$+$	$+$	$+$	$+$	0

$W(-5) - W(5) = 3 - 0 = 3$. The polynomial $f(x)$ has three distinct roots in the segment $[-5, 5]$.

The following simplifications are possible in constructing a Sturm series:

(1) if $f'(x)$ or one of the remainders $r_i(x)$ $(1 \leqslant i < m - 1)$ does not have real roots (does not have roots in the segment $[a, b]$), we can break off the construction of the Sturm series at this point ([23], p. 143);

(2) in carrying out division, we can multiply the divisor and dividend by arbitrary positive numbers (the same);

(3) if $f'(x)$ (or one of the polynomials $f_i(x)$, $1 < i \leqslant m - 1$) is divisible without remainder by a polynomial $g(x)$ which takes non-negative values for all real values of x and reduces to zero only for values $x = \alpha$ for which $f(\alpha) \neq 0$, then $f'(x)$ (or respectively $f_i(x)$) can be divided by $g(x)$ and in the remainder of the construction of the Sturm series we can take the quotient obtained instead of $f'(x)$ (instead of $f_i(x)$).

EXAMPLE 11. $f(x) = x^6 + 4x^5 - 3x^2 + 12x - 24$. Here $f'(x) = 6(x^5 + 2x^2 - x + 2)$ has the number i as a root and therefore is divisible by $x^2 + 1$ (section 3, subsection 6). After cancelling by this polynomial and by the number 6 we obtain $f_1(x) = x^3 - x + 2$. We divide $f(x)$ by $f_1(x)$ and we take the remainder after this division (cancelled by 2) with the opposite sign $f_2(x) = x^2 - 6x + 14$. This polynomial does not have real roots (section 3, subsection 1) and we can stop at this point. We have obtained a Sturm series for the polynomial $f(x)$.

If we need to determine the number of all the distinct real roots of the polynomial $f(x)$ with real coefficients, then we can

apply Sturm's theorem to $f(x)$ in the segment $[a, b]$, where a is zero, and b is an upper bound of the real roots of $f(x)$ (subsection 1). A simpler method is to apply Sturm's theorem in the segment $[a, b]$, where a is a negative number large enough in absolute value, and b is a positive number large enough so that when $x = a$ and when $x = b$ the signs of all the terms of the Sturm series constructed for the polynomial $f(x)$ coincide with the signs of the leading coefficients of these polynomials; such numbers a and b always exist; the symbolic notation for them is $-\infty$ and $+\infty$; in practice it is not necessary to find them explicitly.

If we need to know separately how many distinct positive and how many distinct negative roots the polynomial $f(x)$ has, then we apply Sturm's theorem to the segments $[-\infty, 0]$ and $[0, +\infty]$. If we need to know more precisely in what intervals the roots lie, then we consider the signs of the values of the polynomials of the Sturm series at separate points of these segments.

EXAMPLE 12. $f(x) = 4x^6 + 4x^5 - 15x^4 - 16x^3 + 12x^2 + 16x + 4$.

We take $f_1(x) = \frac{1}{4} f'(x) = 6x^5 + 5x^4 - 15x^3 - 12x^2 + 6x + 4$,

$$f_2(x) = 50x^4 + 57x^3 - 84x^2 - 114x - 32$$

(this is the remainder taken with the opposite sign after the division of the polynomial $9f(x)$ by $f_1(x)$), $f_3(x) = 2x^3 + x^2 - 4x - 2$ (the remainder with the opposite sign after the division of $f_1(x)$ by $f_2(x)$, multiplied by 625/882). $f_2(x)$ is divisible by $f_3(x)$ without remainder. We form the table:

x	$f(x)$	$f_1(x)$	$f_2(x)$	$f_3(x)$	W
$-\infty$	+	−	+	−	3
$+\infty$	+	+	+	+	0
0	+	+	−	−	1
−1	+	0	−	+	2
1	+	−	−	−	1
2	+	+	+	+	0

$W(-\infty) - W(+\infty) = 3$, i.e. $f(x)$ has three distinct real roots.
$W(-\infty) - W(0) = 2$, $W(+\infty) = 1$, i.e. two of them are negative and one is positive.

$$W(-\infty) - W(-1) = 1, \quad W(-1) - W(0) = 1,$$

i.e. one of the negative roots is less than -1, the other is included between -1 and 0. $W(1) - W(2) = 1$, i.e. the positive root lies in the segment $[1, 2]$.

If we need to determine how many roots, each of a given multiplicity, (section 1, subsection 10) the polynomial $f(x)$

has in the interval $[a, b]$, then we have to construct the polynomials $d_1(x) = (f(x), f'(x))$, $d_2(x) = (d_1(x), d_1'(x))$ and so on up to $d_s(x) = (d_{s-1}(x), d_{s-1}'(x))$ (section 1, subsection 5) and for each of them to determine how many distinct roots this polynomial has in the segment $[a, b]$. If $f(x)$ has k_0 distinct roots in the segment $[a, b]$, and $d_i(x)$ has k_i distinct roots $(i = 1, 2, \ldots, s)$, then $k_0 - k_1$ simple roots of $f(x)$ lie in the segment $[a, b]$, $k_1 - k_2$ double roots, $k_2 - k_3$ triple roots and so on.

EXAMPLE 13. Let $f(x)$ be the polynomial considered in the previous example. For it $d_1(x) = f_3(x) = 2x^3 + x^2 - 4x - 2$, $d_2(x) = 1$. By Sturm's method we find that in each of the segment $[-\infty, -1]$, $[-1, 0]$, $[1, 2]$, $d_1(x)$ has one root. The polynomial $d_2(x)$ has no roots. Therefore, in each of the segments considered $f(x)$ has one double root.

The method of quadratic forms. We have the polynomial

$$f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$$

with real coefficients. We construct the quadratic form (Chapter II, section 2, subsection 1)

$$\varphi_0(x_0, \ldots, x_{n-1}) = \sum_{i,k=0}^{n-1} s_{i+k}x_ix_k,$$

where s_{i+k} is the sum of the $(i + k)$th powers of the roots of this polynomial (every root is taken as many times as its multiplicity).† The rank of this quadratic form (Chapter II, section 2, subsection 1) is equal to the number of distinct roots of the polynomial $f(x)$ (real and complex together), and its signature (Chapter II, section 2, subsection 2) to the number of distinct real roots of $f(x)$ ([6], p. 442).

EXAMPLE 14.

$$f(x) = x^5 - 2x^2 + 5x - 8.$$

Here

$$s_0 = 3, \quad s_1 = 2, \quad s_2 = -6, \quad s_3 = 2, \quad s_4 = 50;$$

$$\varphi_0(x_0, x_1, x_2) = 3x_0^2 + 4x_0x_1 - 12x_0x_2 - 6x_1^2 + 4x_1x_2 + 50x_2^2.$$

The rank of the quadratic form φ_0 is 3, the signature is 1 (for the method of determining the rank and signature see Chapter II, section 2, subsections 1 and 2). Hence all three roots of the polynomial $f(x)$ are distinct, and only one of them is real.

---

† The coefficients s_{i+k} can be found by using their expressions in terms of the coefficients of the polynomial $f(x)$ (section 5, subsection 3).

If we need to determine separately the number of all the positive and all the negative roots of $f(x)$, then the result mentioned above can be supplemented with the theorem: the number of distinct pairs of complex conjugate roots of the polynomial $f(x)$ plus the number of distinct negative roots of this polynomial is equal to the number of negative squares in the diagonal form (p. 120) of the quadratic form

$$\varphi_1(x_0, \ldots, x_{n-1}) = \sum_{i,k=0}^{n-1} s_{1+i+k} x_i x_k$$

([15], p. 161).

EXAMPLE 15. $f(x) = 2x^4 - 2x^2 - 2x + 1$. Here

$$s_0 = 4, \quad s_1 = 0, \quad s_2 = 2, \quad s_3 = 3, \quad s_4 = 0, \quad s_5 = 5, \quad s_6 = 2, \quad s_7 = \frac{7}{2};$$

$$\varphi_0(x_0, x_1, x_2, x_3) = 4x_0^2 + 4x_0x_2 + 6x_0x_3 + 2x_1^2 + 6x_1x_2 + 10x_2x_3 + 2x_3^2;$$

$$\varphi_1(x_0, x_1, x_2, x_3) = 4x_0x_1 + 6x_0x_2 + 3x_1^2 + 10x_1x_3 + 5x_2^2 + 4x_2x_3 + \frac{7}{2}x_3^2 = z_0^2 + z_1^2 + z_2^2 - z_3^2$$

(the diagonal form of φ_1). The rank of the quadratic form φ_0 is 4, the signature is 2, therefore $f(x)$ has two distinct real and two distinct complex roots (conjugate to each other (section 3, subsection 6)). In the diagonal form of φ_1 there is one negative square. Because we know already that $f(x)$ has one pair of complex conjugate roots, both the real roots of $f(x)$ are positive.

Note. If

$$f(x) = (a_0 + b_0 i)\, x^n + (a_1 + b_1 i)\, x^{n-1} + \ldots + (a_{n-1} + b_{n-1} i)\, x + (a_n + b_n i)$$

is a polynomial with complex coefficients (a_k, b_k, $k = 0, 1, \ldots, n$, are real numbers), then it can be represented in the form $f(x) = g(x) + h(x)\, i$, where $g(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n$ and $h(x) = b_0 x^n + b_1 x^{n-1} + \ldots + b_{n-1} x + b_n$ are polynomials with real coefficients. The real roots of the polynomial $f(x)$ are all the real roots of the greatest common divisor of the polynomials $g(x)$ and $h(x)$ and nothing else.

3. Bounds on the complex roots of a polynomial.

If

$$f(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n \tag{3.27}$$

is a polynomial with real or complex coefficients, then all its roots can be represented as points on the complex plane ([11],

p. 115) lying inside the circle with centre at the origin of coordinates and radius $N = 1 + A/|a_0|$, where A is the largest of the numbers $|a_1|, \ldots, |a_n|$.

EXAMPLE 1.

$$f(x) = (1 + i)\,x^3 - 3x^2 + ix + 2.$$

Here $A = 3$, $N = 1 + 3/\sqrt{2} \approx 3{\cdot}122$.

The number N can be replaced by the number $N_1 = \rho + A_1/|a_0|$, where ρ is any positive number, and A_1 is the largest of the numbers

$$|a_1|, \quad \frac{|a_2|}{\rho}, \quad \frac{|a_3|}{\rho^2}, \ldots, \quad \frac{|a_n|}{\rho^{n-1}}$$

([24], Problem 686).

EXAMPLE 2.

$$f(x) = x^4 - ix^3 + (3 - 4i)\,x + 16.$$

Here $N = 17$. We take $\rho = 2$. Then we obtain $A_1 = 2$ (as the largest of the numbers 1, 0/2, 5/4, 16/8), $N_1 = 2 + 2 = 4$. All the roots of $f(x)$ lie inside the circle with centre at the origin of coordinates and radius 4.

For other formulae for numbers which can replace N, see [24], Chapter V, section 7.

If all the coefficients of the polynomial (3.27) are real and positive, then under the conditions $a_0 > a_1 > \ldots > a_n$ all its roots lie inside the circle of radius 1 with centre at the origin of coordinates, and under the conditions $a_0 < a_1 < \ldots < a_n$ all the roots lie outside this circle. A more general result is: if all the coefficients of the polynomial (3.27) are strictly positive and if all the ratios a_k/a_{k-1} of one coefficient over the one preceding it are contained between the positive numbers m and M $(m < a_k/a_{k-1} < M, k = 1, 2, \ldots, n)$, then all the roots of this polynomial lie in the annulus lying between the circles with centres at the origin of coordinates and radii m and M respectively ([11], 4th edition, pp. 295–6).

EXAMPLE 3.

$$f(x) = 11x^5 + 8x^4 + 7x^3 + 5x^2 + 2x + 1.$$

The coefficients are positive and decrease, all the roots of $f(x)$ lie inside the circle with radius 1 with centre at the origin of coordinates.

EXAMPLE 4.

$$f(x) = 3x^4 + 3x^3 + 5x^2 + 2x + 4.$$

The ratios of the coefficients are: 1, 5/3, 2/5, 2. All the roots of $f(x)$ can be represented as points on the complex plane lying between the circles of radii 2/5 and 2 with centres at the origin of coordinates (or on these circles themselves).

On the problem of determining the number of roots of a polynomial with real coefficients lying inside or outside an arbitrary given circle, see [15], section 10 or [29], Chapter I, section 8.

If the coefficients a_0, a_1 of the polynomial (3.27) are real, where $a_0 \geqslant 1, a_1 \geqslant 0$, then for the real part $\mathrm{Re}(\alpha)$ of every root α of the polynomial (3.27) the inequality $\mathrm{Re}(\alpha) < (1 + \sqrt{(4c+1)})/2$ holds, where c is any number such that $|a_i| \leqslant c$, $i = 2, 3, \ldots, n$ ([31], pp. 114–7). If the coefficients a_0, a_1 of the polynomial (3.27) do not satisfy the given conditions, then we must make the substitution $y = x + a_1/na_0$, giving a polynomial in which there is no term in the $(n-1)$th power of the unknown, and then multiply the polynomial obtained by a number such that we obtain a polynomial $h(y) = b_0 y^n + b_2 y^{n-2} + \ldots + b_{n-1} y + b_n$ with real leading coefficient $b_0 \geqslant 1$. The real part $\mathrm{Re}(\beta)$ of every root β of the polynomial $h(y)$ satisfies the inequalities

$$-\frac{1+\sqrt{(4c_1+1)}}{2} < \mathrm{Re}(\beta) < \frac{1+\sqrt{(4c_1+1)}}{2},$$

where c_1 is a number such that $|b_i| \leqslant c_1$, $i = 2, 3, \ldots, n$, and for the real parts $\mathrm{Re}(\alpha)$ of the roots α of the polynomial (3.27) the inequalities

$$-\frac{1+\sqrt{(4c_1+1)}}{2} - \mathrm{Re}\left(\frac{a_1}{na_0}\right) < \mathrm{Re}(\alpha) < \frac{1+\sqrt{(4c_1+1)}}{2} - \mathrm{Re}\left(\frac{a_1}{na_0}\right) \quad (3.28)$$

hold ($\mathrm{Re}(a_1/na_0)$ is the real part of a_1/na_0). Geometrically, this means that all the roots of the polynomial (3.27) can be represented as points on a complex plane, lying on the strip bounded by straight lines parallel to the imaginary axis and passing through the points representing the numbers on the left- and right-hand sides of the inequalities (3.28).

EXAMPLE 5.

$$f(x) = (1+i)x^3 - 3x^2 + ix + 2.$$

The substitution $y = x - 1/(1+i)$ gives the polynomial

$$g(y) = (1+i)y^3 + \frac{-4+i}{1+i}y + \frac{-3+5i}{2i}.$$

We multiply this polynomial by $1/(1+i)$ and we obtain

$$h(y) = y^3 + \frac{1+4i}{2}y + \frac{4-i}{2}.$$

We take

$$c_1 = 2,\ 1 \left(> \left| \frac{1+4i}{2} \right| = \left| \frac{4-i}{2} \right| = \frac{\sqrt{17}}{2} \right).$$

Then

$$\frac{1+\sqrt{(4c_1+1)}}{2} = \frac{1+\sqrt{(8\cdot 4+1)}}{2} < \frac{1+3\cdot 1}{2} = 2\cdot 05\,.$$

Because

$$\mathrm{Re}\left(\frac{a_1}{3a_0}\right) = \mathrm{Re}\left(\frac{-1}{1+i}\right) = -\frac{1}{2},$$

for the real parts of the roots α of the polynomial $f(x)$ we obtain $-1\cdot 55 < \mathrm{Re}(\alpha) < 2\cdot 55$.

If we need to establish bounds for the imaginary parts of the roots of the polynomial $f(x)$, then we can make the substitution $x = iy$ in this polynomial. If the real parts of the roots of the polynomial $g(y) = f(iy)$ are contained between the numbers c and d, then the coefficients of the imaginary parts of all the roots of the polynomial $f(x)$ are contained between the same numbers.

EXAMPLE 6.

$$f(x) = ix^5 - (1+i)x^3 - ix^2 + (2-i)x + 6.$$

The substitution $x = iy$ gives

$$g(y) = -y^5 + (-1+i)y^3 + iy^2 + (1+2i)y + 6.$$

For the real parts $\mathrm{Re}(\beta)$ of the roots of the polynomial $g(y)$ the inequalities

$$-\frac{1+\sqrt{((4\times 6)+1)}}{2} = -3 < \mathrm{Re}(\beta) < 3 = \frac{1+\sqrt{(4\times 6+1)}}{2}$$

hold. Hence, all the roots of the polynomial $f(x)$ can be represented as points on the complex plane lying in the strip bounded by staright lines parallel to the real axis and passing through the points representing the numbers $-3i$ and $3i$.

4. The number of roots in the left and right half-planes.

(a) *Cauchy's indices* ([6], Chapter XV, sections 2 and 3).

We have the rational function

$$\Phi(x) = \frac{k(x)}{h(x)},$$

where $k(x)$ and $h(x)$ are polynomials with real coefficients. The *Cauchy index* of the function $\Phi(x)$ within the limits from a to b (a and b are real numbers) is the difference between the number of discontinuities in $\Phi(x)$ ([34], p. 48) in the passage

from $-\infty$ to $+\infty$ and the number of discontinuities in the passage from $+\infty$ to $-\infty$, obtained when x varies from a to b. The index is denoted by $I_a^b \Phi(x)$. If $\Phi(x)$ has a discontinuity when $x = a$ or when $x = b$, we do not take this discontinuity into account.

EXAMPLE 1.

$$\Phi(x) = \frac{2x+1}{(x+2)^3(x-1)(x-4)(x-5)^2}.$$

If here x, as it increases, passes through the points $x = -2$ or $x = 1$, then the sign of $\Phi(x)$ changes from $+$ to $-$; if x passes through the point $x = 4$, then the sign changes from $-$ to $+$; in passing through the point $x = 5$ the sign of $\Phi(x)$ does not change. Therefore, for instance, $I_{-3}^{6}\Phi(x) = 1 - 2 = -1$. If we make x vary from 6 to -3, then we obtain $I_{6}^{-3}\Phi(x) = 2 - 1 = 1$.

In determining the Cauchy index, instead of the real numbers a and b we can take $\pm\infty$. In the preceding example

$$I_{-\infty}^{+\infty}\Phi(x) = -1, \quad I_{+\infty}^{-\infty}\Phi(x) = 1.$$

If all the real roots of the polynomial $h(x)$ are known, where $h(x) = (-\alpha_1)^{l_1} \dots (x-\alpha_m)^{l_m} h_1(x)$ ($\alpha_1, \dots, \alpha_m$ are all the distinct real roots of $h(x)$, and $h_1(x)$ is a polynomial without real roots), then the index of the function $\Phi(x) = k(x)/h(x)$ is most easily calculated by representing it in the form

$$\Phi(x) = \sum_{i=1}^{m}\left(\frac{A_1^{(i)}}{x-\alpha_i} + \dots + \frac{A_{l_i}^{(i)}}{(x-\alpha_i)^{l_i}}\right) + \frac{q(x)}{h_1(x)}, \tag{3.29}$$

where the coefficients $A_j^{(i)}$ and the coefficients of the polynomial $q(x)$ are real numbers, and $A_{l_i}^{(i)} \neq 0$, $i = 1, 2, \dots, m$ (such a representation of the function $\Phi(x)$ is always possible); if we denote by $\overline{A}_{l_i}^{(i)}$ those of the coefficients $A_{l_i}^{(i)}$, $i = 1, 2, \dots, m$, which have odd exponents l_i, then we will have

$$I_{-\infty}^{+\infty}\Phi(x) = \sum \operatorname{sign} \overline{A}_{l_i}^{(i)}, \tag{3.30}$$

where $\operatorname{sign} \overline{A}_{l_i}^{(i)}$ is 1 or -1 depending on whether $\overline{A}_{l_i}^{(i)} > 0$ or $\overline{A}_{l_i}^{(i)} < 0$. In practice, when finding $\overline{A}_{l_i}^{(i)}$ we first write equation (3.29) with non-negative coefficients, then reduce all the fractions on the right-hand side of this equation to a common denominator and find $\overline{A}_{l_i}^{(i)}$, starting by equating the numer-

ator obtained in this way on the right-hand side of (3.29) and the numerator of $\Phi(x)$. If we need to determine $I_a^b \Phi(x)$, then in the formula (3.30) we only take those $\overline{A_i^{(i)}}$ for which there exist roots α_i lying between a and b $(a < \alpha_i < b)$.

EXAMPLE 2.

$$\Phi(x) = \frac{x^3 + 1}{x^3(x-2)^2(x+3)(x^4+3x^2+5)}.$$

We represent $\Phi(x)$ in the form (3.29):

$$\Phi(x) = \frac{A_1^{(1)}}{x} + \frac{A_2^{(1)}}{x^2} + \frac{A_3^{(1)}}{x^3} + \frac{A_1^{(2)}}{x-2} + \frac{A_2^{(2)}}{(x-2)^2} + \frac{A_1^{(3)}}{x+3} + \frac{q(x)}{x^4+3x^2+5}$$

(the polynomial $x^4 + 3x^2 + 5$ does not have any real roots). We reduce all the fractions to a common denominator, equate the numerators on the left- and right-hand sides of the equation and first set $x = 0$, and then $x = -3$. This gives

$$A_3^{(1)} = \frac{1}{4 \times 3 \times 5} > 0, \quad A_1^{(3)} = \frac{-27+1}{(-27) \times 25 \times 113} > 0.$$

Hence, $I_{-\infty}^{+\infty} \Phi(x) = 2$.

For other methods of determining the Cauchy index of a rational fraction $\Phi(x)$, see [6], Chapter XV, sections 2 and 11.

Now let $f(x)$ be a polynomial of degree n with real coefficients. We write it in the form

$$f(x) = a_0 x^n + b_0 x^{n-1} + a_1 x^{n-2} + b_1 x^{n-3} + \ldots \tag{3.31}$$

and we set

$$f_1(x) = a_0 x^n - a_1 x^{n-2} + a_2 x^{n-4} - \ldots,$$

$$f_2(x) = b_0 x^{n-1} - b_1 x^{n-3} + b_2 x^{n-5} - \ldots \tag{3.32}$$

Then if t is the number of roots of the polynomial $f(x)$ lying in the right half-plane (i.e. having positive real part—see [11], p. 115), and s is the number of its purely imaginary roots, then

$$I_{-\infty}^{+\infty} \frac{f_2(x)}{f_1(x)} = n - 2t - s$$

(every root is counted as many times as its multiplicity, see section 1, subsection 10).

The number of purely imaginary roots of the polynomial $f(x)$ is equal to the number of real roots of the greatest common divisor of the polynomials (3.32).

EXAMPLE 3.

$$f(x) = x^4 + x^3 - 2x - 1.$$

Here the polynomials $f_1(x) = x^4 - 1$ and $f_2(x) = x^3 + 2x$ are mutually prime (section 1, subsection 6), so that $s = 0$.

$$I_{-\infty}^{+\infty} \frac{x^3 + 2x}{x^4 - 1} = 1 + 1 = 2 = 4 - 2t, \quad t = 1,$$

(b) *Routh's algorithm* ([6], Chapter XV, sections 3 and 4). For the polynomial (3.31) we form Routh's scheme

$$\begin{array}{llll} a_0, & a_1, & a_2, & \ldots, \\ b_0, & b_1, & b_2, & \ldots, \\ c_0, & c_1, & c_2, & \ldots, \\ d_0, & d_1, & d_2, & \ldots, \\ \ldots\ldots\ldots\ldots, & & & \end{array}$$

in the first two rows of which we write the coefficients of the polynomial (3.31), and each of the following rows is obtained from the two preceding ones thus: from a number in the first row we subtract the number below it in the second row, multiplied by a number such that the first difference would be equal to zero; this first difference is rejected, and the row begins with the second difference. It is possible to construct such a scheme if there do not occur any numbers equal to zero in the series $b_0, c_0, d_0, \ldots$, (the so-called *regular case*). In this case, the number of roots of the polynomial (3.31) lying in the right half-plane is equal to the number of changes of sign† in the first column of the scheme constructed (every root is counted as many times as its multiplicity). In the regular case the polynomial $f(x)$ does not have purely imaginary roots.

EXAMPLE 4.

$$f(x) = x^6 - 5x^5 + 2x^4 - x^2 + x + 3.$$

We construct Routh's scheme:

$$\begin{array}{rrrr} 1 & 2 & -1 & 3 \\ -5 & 0 & 1 & \\ 2 & -\frac{4}{5} & 3 & \\ -2 & \frac{17}{2} & & \\ \frac{77}{10} & 3 & & \\ \frac{1429}{154} & & & \\ 3 & & & \end{array}$$

† See footnote on p. 193.

In the first column of this scheme there are four changes of sign, i.e. $f(x)$ has four roots with positive real parts. The remaining two roots of the polynomial $f(x)$ have negative real parts.

If in the construction of the Routh scheme for the polynomial (3.31) there occurs a number $h_0 = 0$ in the series b_0, c_0, d_0, ..., and the whole row containing h_0 turns out to consist of zeros, then we must replace this row by the row of numbers $(n-m+1)e_0$, $(n-m-1)e_1$, $(n-m-3)e_2$, ..., where e_0, e_1, e_2, ... are the elements of the preceding row, and m is the number of rows of the scheme already constructed, and we continue to construct the Routh scheme. If this construction comes to an end (perhaps, after several rows of zeros have been obtained and replaced by the corresponding numbers), then again the number of changes of sign in the first column of the scheme obtained will be equal to the number of roots of the polynomial (3.31) lying in the right half-plane. The number of purely imaginary roots of this polynomial (each counted as many times as its multiplicity) in the given case is equal to $n-m+1-2l$, where m is the number of rows obtained up to the appearance of the first row of zeros, and l is the number of changes of sign in the first column of the scheme constructed beginning with the mth position.

EXAMPLE 5.

$$f(x) = x^8 - x^7 + 2x^6 - x^5 + x^3 - 2x^2 + x - 1.$$

We construct the Routh scheme:

1	2	0	−2	−1	
−1	−1	1	1		
1	1	−1	−1		
6	4	−2			(the row of zeros replaced by the corresponding numbers)
$\frac{1}{3}$	$-\frac{2}{3}$	−1			
16	16				
−1	−1				
−2					(again replacing the row of zeros)
−1					

In the first column of the scheme there are three changes of sign, i.e. three roots of the polynomial $f(x)$ lie in the right half-plane. The number of purely imaginary roots of $f(x)$ is $s = 8 - 3 + 1 - 2 \times 1 = 4$.

EXAMPLE 6.

$$f(x) = 3x^5 + x^4 + 6x^3 + 2x^2 + 3x + 1.$$

We construct the Routh scheme:

3	6	3	
1	2	1	(replacing a row of zeros)
4	4		
1	1		
2			(again replacing a row of zeros)
1			

In the first column of the scheme there are no changes of sign, i.e. $f(x)$ does not have any roots in the right half-plane. The number of purely imaginary roots is $s = 5 - 2 + 1 - 2 \times 0 = 4$.

If we have obtained $h_0 = 0$, but in the row in which h_0 occurs there are also numbers distinct from zero, then we must divide the polynomial (3.31) by the greatest common divisor $D(x)$ of the polynomials $F_1(x) = a_0x^n + a_1x^{n-2} + a_2x^{n-4} + \ldots$ and $F_2(x) = b_0x^{n-1} + b_1x^{n-3} + b_2x^{n-5} + \ldots$ (section 1, subsection 5). If

$$f(x) = D(x)f^*(x), \tag{3.33}$$

then the number of roots of $f(x)$ lying in the right half-plane is equal to $t_1 + t_2$, where t_1 (respectively t_2) is the number of the roots of the polynomial $D(x)$ (respectively of $f^*(x)$) lying in the right half-plane. Here $t_1 = (q - s)/2$, where q is the degree of $D(x)$, s is the number of real roots of the polynomial $D(iy)$ (for the method of finding s see subsection 2), and t_2 is determined by Routh's algorithm, applied to the polynomial $f^*(x)$. If in constructing the Routh scheme for $f^*(x)$ we obtain the number $h_0 = 0$ in the first column, and in the row in which it occurs there are also numbers distinct from zero, then we must replace h_0 by a small quantity ε of definite (but arbitrary) sign and continue to construct the scheme. The signs of the elements obtained subsequently will be determined, depending on the size and sign of ε. If one of the elements of the first column of the scheme turns out to be identically zero with respect to ε, then we must replace it by another small quantity η and continue the calculation. The number of changes of sign in the first column of the scheme constructed will be equal to t_2. The number of purely imaginary roots of the polynomial $f(x)$ in the case under consideration is equal to s.

EXAMPLE 7.

$$f(x) = x^6 - x^4 + 3x^3 - x^2 + 3x + 1.$$

Here $b_0 = 0$. The greatest common divisor of the polynomials $F_1(x) = x^6 - x^4 - x^2 + 1$ and $F_2(x) = 3x^3 + 3x$ is $D(x) = x^2 + 1$, $f(x) = D(x)f^*(x)$,

where $f^*(x) = x^4 - 2x^2 + 3x + 1$; $q = 2$, $s = 2$, $t_1 = 0$. Applying Routh's algorithm to the polynomial $f^*(x)$ (where we introduce the small quantity ε):

$$\begin{array}{cccc} 1 & & -2 & 1 \\ \varepsilon & & 3 & \\ -2 & -\dfrac{3}{\varepsilon} & 1 & \\ 3 & +\dfrac{\varepsilon^2}{2\varepsilon+3} & & \\ 1 & & & \end{array}$$

For any quantity ε sufficiently small in absolute value there are two changes of sign in the first column of this scheme, $t_2 = 2$. The polynomial $f(x)$ has two roots in the right half-plane and two purely imaginary roots.

(c) *The Routh–Hurwitz theorem* ([6], Chapter XV, sections 6, 8 and 12). For the polynomial (3.31) we construct the *Hurwitz matrix*

$$H = \begin{pmatrix} b_0 & b_1 & b_2 & \dots & b_{n-1} \\ a_0 & a_1 & a_2 & \dots & a_{n-1} \\ 0 & b_0 & b_1 & \dots & b_{n-2} \\ 0 & a_0 & a_1 & \dots & a_{n-2} \\ 0 & 0 & b_0 & \dots & b_{n-3} \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

(a square matrix of the nth order [see Introduction], in which the elements a_i, b_i, which do not occur in the polynomial (3.31) are equal to zero). The minors of this matrix of the orders $1, 2, \dots, n$ (p. 31), in the upper left-hand corner, are called the *Hurwitz determinants* and are denoted by

$$\Delta_1, \Delta_2, \dots, \Delta_n. \tag{3.34}$$

If all these minors are distinct from zero, then the number t of roots of the polynomial (3.31) lying in the right half-plane is equal to the number of changes of sign in the series

$$a_0, \Delta_1, \frac{\Delta_2}{\Delta_1}, \frac{\Delta_3}{\Delta_2}, \dots, \frac{\Delta_n}{\Delta_{n-1}}. \tag{3.35}$$

The polynomial (3.31) does not have any purely imaginary roots in this case.

EXAMPLE 8.

$$f(x) = 3x^5 + x^4 + 4x^3 + 2x^2 + 6x + 2.$$

We form the Hurwitz matrix:

$$H = \begin{pmatrix} 1 & 2 & 2 & 0 & 0 \\ 3 & 4 & 6 & 0 & 0 \\ 0 & 1 & 2 & 2 & 0 \\ 0 & 3 & 4 & 6 & 0 \\ 0 & 0 & 1 & 2 & 2 \end{pmatrix}$$

$a_0 = 3$, $\Delta_1 = 1$, $\Delta_2 = -2$, $\Delta_3 = -4$, $\Delta_4 = -8$, $\Delta_5 = -16$. In the series $3, 1, -2/1 = -2, -4/-2 = 2, -8/-4 = 2, -16/-8 = 2$ there are two changes of sign, $f(x)$ has 2 roots in the right half-plane.

If several of the minors (3.34) are equal to zero, but $\Delta_n \neq 0$, then necessarily $\Delta_{n-1} \neq 0$, and if

$$\Delta_s \neq 0, \quad \Delta_{s+1} = \ldots = \Delta_{s+p} = 0, \quad \Delta_{s+p+1} \neq 0, \tag{3.36}$$

then $\Delta_{s-1} \neq 0$ (if $s > 1$), $\Delta_{s+p+2} \neq 0$ and p is an odd number. The number t of roots of the polynomial (3.31) lying in the right half-plane is equal to the number of changes of sign in the series (3.35), where we only need consider that under the conditions (3.36) between the numbers Δ_s/Δ_{s-1} and $\Delta_{s+p+2}/\Delta_{s+p+1}$ there are $(p+2)/2 - (-1)^{(p+1)/2}\,\delta/2$ changes of sign (δ is either 1 or -1 depending on whether Δ_s/Δ_{s-1} and $\Delta_{s+p+2}/\Delta_{s+p+1}$ have the same or opposite signs); when $s = 1$, Δ_s/Δ_{s-1} is replaced by Δ_1; if $\Delta_1 = \ldots \Delta_p = 0$, $\Delta_{p+1} \neq 0$, then by the same rule we determine the number of changes of sign between the numbers a_0 and $\Delta_{p+2}/\Delta_{p+1}$ (here again necessarily $\Delta_{p+2} \neq 0$ and p is an odd number). When $\Delta_n \neq 0$ the polynomial (3.31) does not have any purely imaginary roots.

EXAMPLE 9.

$$f(x) = x^3 + x + 2.$$

Here

$$\Delta_3 = \begin{vmatrix} 0 & 2 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{vmatrix} = -4, \quad \Delta_2 = -2.$$

$\Delta_1 = 0$, $a_0 = 1$; p = 1. The number of changes of sign between a_0 and Δ_3/Δ_2 is: $t = 3/2 - (-1)\,1/2 = 2$, $f(x)$ has two roots in the right half-plane.

If $\Delta_n = 0$, then we must represent the polynomial (3.31) in the form (3.33) and to determine the number t_2 of roots of the polynomial $f^*(x)$ lying in the right half-plane, we construct the Hurwitz determinants for this polynomial (the one of them whose order is equal to the degree of $f^*(x)$ is necessarily distinct from zero).

It is convenient to use the Routh–Hurwitz theorem when the coefficients of the polynomial depend on parameters

and we are required to determine for what values of the parameters the number t of roots of the polynomial lying in the right half-plane has some particular value.

For some extensions of the Routh–Hurwitz theorem and other methods of determining the number of roots of a polynomial lying in the left or right half-plane, see [6], Chapter XV, and [29], Chapter I (particularly sections 8 and 9).

(d) *The case of complex coefficients* ([6], Chapter XV, section 18). If $f(x)$ is a polynomial with complex coefficients, then to determine the number of roots of $f(x)$ lying in the right half-plane we must first make the substitution $x = iy$. As a result we obtain the polynomial

$$f(iy) = c_0 y^n + c_1 y^{n-1} + \ldots + c_n + i(d_0 y^n + d_1 y^{n-1} + \ldots + d_n),$$

where $c_k, d_k (k = 0, 1, \ldots, n)$ are real numbers. If $d_0 \neq 0$, we must then construct the square matrix of order $2n$:

$$A = \begin{pmatrix} d_0 & d_1 \ldots d_n & 0 & \ldots 0 \\ c_0 & c_1 \ldots c_n & 0 & \ldots 0 \\ 0 & d_0 \ldots d_{n-1} & d_n & \ldots 0 \\ 0 & c_0 \ldots c_{n-1} & c_n & \ldots 0 \\ \ldots & \ldots & \ldots & \ldots \end{pmatrix}$$

(if $d_0 = 0$, then instead of $f(iy)$ we must take the polynomial $if(iy)$ and construct the corresponding matrix for it). The minors of this matrix (p. 31) of even order standing in the upper left-hand corner, will be denoted by $\nabla_2, \nabla_4, \ldots, \nabla_{2n}$. If $\nabla_{2n} \neq 0$, then the number t of roots of the polynomial $f(x)$ lying in the right half-plane is equal to the number of changes of sign in the series $1, \nabla_2, \nabla_4, \ldots, \nabla_{2n}$, where, if there are any zeros in this series, then for every group of successive zeros

$$(\nabla_{2h} \neq 0), \quad \nabla_{2h+2} = \ldots = \nabla_{2h+2p} = 0, \quad (\nabla_{2h+2p+2} \neq 0)$$

we must assume that between ∇_{2h} and $\nabla_{2h+2p+2}$ there are $(p+1)/2$ (changes of sign), when p is odd, and $(p+1-\delta)/2$, when p is even, (δ is equal to 1 or -1 depending on whether the numbers $(-1)^{p/2}$ and $(\nabla_{2h+2p+2})/\nabla_{2h}$ have the same or opposite signs). When $\nabla_{2n} \neq 0$ the polynomial $f(x)$ has no purely imaginary roots.

EXAMPLE 10.

$$f(x) = ix^3 - x^2 + (5 - 11i)x - (4 + 20i).$$

Here

$$if(iy) = iy^3 + iy^2 + (-5 + 11i)y - (-20 + 4i)$$
$$= -5y + 20 + i(y^3 + y^2 + 11y - 4),$$

$$A = \begin{pmatrix} 1 & 1 & 11 & -4 & 0 & 0 \\ 0 & 0 & -5 & 20 & 0 & 0 \\ 0 & 1 & 1 & 11 & -4 & 0 \\ 0 & 0 & 0 & -5 & 20 & 0 \\ 0 & 0 & 1 & 1 & 11 & -4 \\ 0 & 0 & 0 & 0 & -5 & 20 \end{pmatrix}.$$

We write out the series 1, $\nabla_2 = 0$, $\nabla_4 = -25$, $\nabla_6 = -15{,}000$, $p = 1$; we assume that between 1 and ∇_4 we have one change of sign. There are no other changes of sign, $t = 1$.

If $\nabla_{2n} = 0$, then we can find the greatest common divisor $\overline{D}(y)$ (section 1, subsection 5) of the polynomials $\overline{F}_1(y) = c_0 y^n + c_1 y^{n-1} + \ldots + c_n$ and $\overline{F}_2(y) = d_0 y^n + d_1 y^{n-1} + \ldots + d_n$ and divide $f(x)$ by the polynomial $\overline{D}(-ix)$. As the quotient we obtain a polynomial $\overline{f}(x)$ of degree k, for which $\nabla_{2k} \neq 0$. The number t of roots $f(x)$ lying in the right half-plane is equal to the sum of the number $\overline{t}_1$ of roots of $\overline{f}(x)$ lying in the right half-plane (this number is found in the way given above), and the number $\overline{t}_2$ of roots of $\overline{D}(-ix)$ lying in the right half-plane (it is equal to $(\overline{q} - \overline{s})/2$, where $\overline{q}$ is the degree, s the number of real roots (section 4, subsection 2) of the polynomial $\overline{D}(y)$. The number of purely imaginary roots of $f(x)$ in the given case is equal to $\overline{s}$.

EXAMPLE 11.

$$f(x) = x^5 + (-1 + 2i)x^4 + (-1 - i)x^3 + ix^2 + (-2 - i)x + (1 - i).$$

Here

$$f(iy) = iy^5 + (-1 + 2i)y^4 + (-1 + i)y^3 - iy^2 + (1 - 2i)y + (1 - i)$$
$$= (-y^4 - y^3 + y + 1) + i(y^5 + 2y^4 + y^3 - y^2 - 2y - 1);\ \nabla_{10} = 0$$

The greatest common divisor of the polynomials

$$\overline{F}_1(y) = -y^4 - y^3 + y + 1 \quad \text{and} \quad \overline{F}_2(y) = y^5 + 2y^4 + y^3 - y^2 - 2y - 1$$

is

$$\overline{D}(y) = y^4 + y^3 - y - 1.\quad \overline{D}(-ix) = x^4 + ix^3 + ix - 1;$$

$$\overline{f}(x) = x - 1 + i;\quad \overline{t}_1 = 1,\quad \overline{s} = 2,\quad \overline{t}_2 = \frac{4 - 2}{2} = 1.$$

$f(x)$ has two roots in the right half-plane and two purely imaginary roots.

5. Criteria of stability.

In a number of practical fields, where one is studying the stability of mechanical and electrical systems, it is necessary to deal with conditions under which all the roots of a polynomial lie in the left half-plane, i.e. have negative real part ([11], p. 115). A polynomial with this property is called a *Hurwitz polynomial.*

If $f(x)$ is a Hurwitz polynomial with real coefficients, then all its coefficients are necessarily positive ([11], 4th edition, p. 318).

Routh's criterion ([6], p. 425). In order that all the roots of the polynomial $f(x)$ with real coefficients should lie in the left half-plane, it is necessary and sufficient that all the elements of the first column of Routh's scheme (subsection 4 (b)) should be distinct from zero and of one sign.

EXAMPLE 1.

$$f(x) = x^3 + x^2 + 3x + 2.$$

We construct Routh's scheme:

$$\begin{array}{ll} 1 & 3 \\ 1 & 2 \\ 1 & \\ 2 & \end{array}$$

All the elements of the first column of the scheme are positive, $f(x)$ is a Hurwitz polynomial.

EXAMPLE 2. $f(x) = 3x^5 + x^4 + 6x^3 + 2x^2 + 3x + 1$ (subsection 4 (b)). This polynomial is not a Hurwitz polynomial, which can be seen immediately from the fact that in constructing its Routh scheme we obtain a zero in the first column (a row of zeros).

Hurwitz's criterion ([6], p. 436, [11], 4th edition, p. 316). In order that all the roots of the polynomial $f(x)$ with real coefficients and positive leading coefficient should lie in the left half-plane, it is necessary and sufficient that all the Hurwitz determinants (subsection 4 (c)) constructed for this polynomial should be positive.

If the leading coefficient of $f(x)$ is negative, then we must consider the polynomial $-f(x)$ whose roots coincide with the roots of $f(x)$.

EXAMPLE 3.

$$f(x) = x^4 + 2x^3 + 9x^2 + x + 4.$$

We construct the Hurwitz matrix (subsection 4 (c)):

$$H = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 9 & 4 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 9 & 4 \end{pmatrix}.$$

$\Delta_1 = 2 > 0$, $\Delta_2 = 17 > 0$, $\Delta_3 = 1 > 0$, $\Delta_4 = 4 > 0$; $f(x)$ is a Hurwitz polynomial.

The criterion of Liénard and Chipart ([6], p. 457). Necessary and sufficient conditions for all the roots of the polynomial $f(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n$ with real coefficients and $a_0 > 0$ to lie in the left half-plane can be written in any of the following four ways:

(1) $a_n > 0$, $a_{n-2} > 0, \ldots$; $\Delta_1 > 0$, $\Delta_3 > 0, \ldots$;
(2) $a_n > 0$, $a_{n-2} > 0, \ldots$; $\Delta_2 > 0$, $\Delta_4 > 0, \ldots$;
(3) $a_n > 0$, $a_{n-1} > 0$, $a_{n-3} > 0, \ldots$; $\Delta_1 > 0$, $\Delta_3 > 0, \ldots$;
(4) $a_n > 0$, $a_{n-1} > 0$, $a_{n-3} > 0, \ldots$; $\Delta_2 > 0$, $\Delta_4 > 0, \ldots$;

$\Delta_1, \Delta_2, \Delta_3, \ldots$ are the Hurwitz determinants (subsection 4 (c)).

EXAMPLE 4.

$$f(x) = x^6 + x^5 + 3x^4 + x^3 + 5x^2 + 2x + 1.$$

Here all the coefficients are positive. We construct the Hurwitz matrix:

$$H = \begin{pmatrix} 1 & 1 & 2 & 0 & 0 & 0 \\ 1 & 3 & 5 & 1 & 0 & 0 \\ 0 & 1 & 1 & 2 & 0 & 0 \\ 0 & 1 & 3 & 5 & 1 & 0 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 1 & 3 & 5 & 1 \end{pmatrix}$$

$\Delta_1 = 1 > 0$, but $\Delta_3 = -1 < 0$, i.e. not all the roots of the given polynomial have negative real part.

For other criteria of stability see [6], Chapter XV, sections 5, 14 and 15, [15] and [29].

6. The domain of stability.

Every polynomial of the nth degree with real coefficients can be made to correspond to the points of n-dimensional space whose coordinates are equal to the quotients after division of all the remaining coefficients of the polynomial by the leading coefficient. In such a "space of coefficients" the Hurwitz polynomials form an n-dimensional domain, which is called the *domain of stability*. If the coefficients of the polynomial are given as functions of p parameters, then

the domain of stability is constructed in the space of these parameters. For a study of the domain of stability (which is essential, for instance, in the projection of new systems of reference) see [6], Chapter XV, sections 15 and 17.

5. Polynomials in Several Variables

1. Definitions.

A *polynomial* $f(x_1, x_2, \ldots, x_n)$ *in the n variables* $x_1, x_2, \ldots, x_n$ *over the number field* K (see Introduction) is the sum of a finite number of terms of the form $x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}$, where $k_1, k_2, \ldots, k_n$ are non-negative integers, with coefficients from the field K; here it is presupposed that the polynomial does not contain similar terms, and the terms with coefficients equal to zero are not considered. Two polynomials $f_1(x_1, x_2, \ldots, x_n)$ and $f_2(x_1, x_2, \ldots, x_n)$ are called *equal* if their coefficients for similar terms are equal.

The sum $k_1 + k_2 + \ldots + k_n$ is called the *degree of the term* $x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}$ of the polynomial $f(x_1, x_2, \ldots, x_n)$; the largest of these sums is called the *degree of the polynomial* $f(x_1, x_2, \ldots, x_n)$ *with respect to all its variables.* Polynomials of zero degree are elements of K distinct from zero. The number zero is also regarded as a polynomial; this is the only polynomial in n variables whose degree is not defined. If all the terms of the polynomial $f(x_1, x_2, \ldots, x_n)$ have the same degree s with respect to all the variables, then such a polynomial is called a *homogeneous polynomial* or a *form* of the sth degree in the n variables (examples: linear forms, quadratic forms—see Chapter II, section 2, subsection 1).

The *degree* of the polynomial $f(x_1, x_2, \ldots, x_n)$ *with respect to one of the variables* x_i $(i = 1, 2, \ldots, n)$ is the highest exponent with which x_i occurs in the terms of this polynomial (this degree may also be zero).

Of the two terms $x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}$ and $x_1^{l_1} x_2^{l_2} \ldots x_n^{l_n}$ of the polynomial $f(x_1, x_2, \ldots, x_n)$ the one which is considered to be the *higher* is that in which the exponent of x_1 is larger, or if these exponents are equal, then that in which the exponent of x_2 is larger, and so on (i.e. the first term is higher than the second if the first of the differences $k_i - l_i$ $(i = 1, 2, \ldots, n)$ distinct from zero is positive). If all the terms of the polynomial $f(x_1, x_2, \ldots, x_n)$ are arranged in the order in which each term is smaller than the one preceding it, then we say that the terms of this

polynomial are arranged *lexicographically*. The term which stands first here is called the *highest term* of the polynomial.

EXAMPLE.

$$f(x_1, x_2, x_3) = x_1^2x_2 - 2x_1^2x_3 - 8x_1x_2^5x_3^7 + x_2^6\,x_3 + x_2^2$$

is the lexicographic expression for the polynomial, $x_1^2\,x_2$ is the highest term.

The *sum* of the polynomials $f_1(x_1, x_2, \ldots, x_n)$ and $f_2(x_1, x_2, \ldots, x_n)$ is the polynomial whose coefficients are obtained by adding corresponding coefficients of f_1 and f_2; if here some term occurs in only one of the polynomials f_1, f_2, then it is regarded as occurring in the other polynomial with coefficient zero. The *product* of the polynomials $f_1(x_1, x_2, \ldots, x_n)$ and $f_2(x_1, x_2, \ldots, x_n)$ is the polynomial obtained by multiplying f_1 by f_2 term by term and then collecting similar terms. The degree (with respect to all the variables) of the product of two non-zero polynomials in n variables is equal to the sum of the degrees of these polynomials, the highest term of the product is equal to the product of the highest terms of the factors.

2. Symmetric polynomials.

The polynomial $f(x_1, x_2, \ldots, x_n)$ in n variables is called *symmetric* if it remains unchanged under any permutation of these variables. In order to verify whether a given polynomial is symmetric, it is sufficient to verify whether it is altered by permutations of any two of its variables with each other.

EXAMPLE 1. The polynomial

$$f(x_1, x_2, x_3) = x_1^2x_2x_3 + x_1x_2^2x_3 + x_3^2$$

is not symmetric, because, if we replace x_1 by x_3, and x_3 by x_1 everywhere in it, we obtain the polynomial

$$x_3^2x_2x_1 + x_3x_2^2x_1 + x_1^2 \neq f(x_1, x_2, x_3).$$

EXAMPLE 2.

$$f(x_1, x_2, x_3) = x_1^3 + x_2^3 + x_3^3 - 2x_1x_2 - 2x_1x_3 - 2x_2x_3 + 10$$

is a symmetric polynomial in the variables x_1, x_2, x_3.

The symmetric polynomials in the n variables

$$\begin{aligned}
\sigma_1 &= x_1 + x_2 + \ldots + x_n,\\
\sigma_2 &= x_1x_2 + x_1x_3 + \ldots + x_{n-1}x_n,\\
\sigma_3 &= x_1x_2x_3 + x_1x_2x_4 + \ldots + x_{n-2}x_{n-1}x_n,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
\sigma_{n-1} &= x_1x_2\ldots x_{n-1} + x_1x_2\ldots x_{n-2}x_n + \ldots + x_2x_3\ldots x_n,\\
\sigma_n &= x_1x_2\ldots x_n
\end{aligned}$$

are called the *elementary* (or *fundamental*) symmetric polynomials. If $g(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$ is a polynomial with coefficients from the field K in one variable (section 1, subsection 1), then the values of the elementary symmetric polynomials in n variables for values of the variables equal to the roots of the polynomial $g(x)$, are equal respectively to $-\frac{a_1}{a_0}, \frac{a_2}{a_0}, \ldots, (-1)^n\frac{a_n}{a_0}$ (cf. section 1, subsection 11).

A symmetric polynomial in the variables $x_1, x_2, \ldots, x_n$, all of whose terms can be obtained from one of them by permuting the variables $x_1, x_2, \ldots, x_n$, is called a *monogenous* polynomial ([23], p. 214). If $ax_1^{k_1}x_2^{k_2} \ldots x_n^{k_n}$ is the highest term of a monogenous polynomial, then this polynomial is denoted by $S(ax_1^{k_1}x_2^{k_2} \ldots x_n^{k_n})$.

EXAMPLE 3. A monogenous polynomial in three variables is

$$S(x_1^2x_2) = x_1^2x_2 + x_1^2x_3 + x_2^2x_1 + x_2^2x_3 + x_3^2x_1 + x_3^2x_2.$$

EXAMPLE 4.

$$f(x_1, x_2, x_3) = x_1^3 + x_2^3 + x_3^3 - 2x_1x_2 - 2x_1x_3 - 2x_2x_3 + 10 = S(x_1^3) + S(-2x_1x_2) + 10.$$

The fundamental theorem on symmetric polynomials. Every symmetric polynomial in the variables $x_1, x_2, \ldots, x_n$ can be represented in the form of a polynomial in the elementary symmetric polynomials (with coefficients from the same field as that in which the coefficients of the symmetric polynomial lie); this representation is unique.

EXAMPLE 5.

$$f(x_1, x_2) = x_1^2 + x_2^2 - x_1x_2 + 3 = \sigma_1^2 - 3\sigma_2 + 3.$$

In order to find an expression for a given symmetric polynomial in terms of the elementary ones, we must first separate this polynomial into homogeneous parts (subsection 1), collecting together all terms of the polynomial which have the same degree with respect to the whole set of variables, and then we must express each of the homogeneous parts separately in terms of the elementary symmetric polynomials. In order to express the homogeneous symmetric polynomial $f(x_1, x_2, \ldots, x_n)$ in terms of the elementary symmetric polynomials, we must take its highest term $a_0x_1^{k_1}x_2^{k_2} \ldots x_n^{k_n}$ (subsection 1), write out the set of its exponents $k_1, k_2, \ldots, k_n$ and form all possible sets of numbers of the form $l_1, l_2, \ldots, l_n$ with the properties: (1) the sum of the numbers $l_1 + l_2 + \ldots + l_n$ in every set is the same and equal to $k_1 - k_2 +, \ldots, + k_n$;

(2) the numbers of each set do not go in ascending order, i.e. $l_1 \geqslant l_2 \geqslant \ldots \geqslant l_n$; (3) the term $x_1^{l_1} x_2^{l_2} \ldots x_n^{l_n}$ is not higher than the term $x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}$. After this for every set $l_1, l_2, \ldots, l_n$ we must form the product $\sigma_1^{l_1-l_2} \sigma_2^{l_2-l_3} \ldots \sigma_{n-1}^{l_{n-1}-l_n} \sigma_n^{l_n}$ and set the polynomial $f(x_1, x_2, \ldots, x_n)$ equal to the sum of the products so constructed, taken with non-negative coefficients.† If we find these coefficients, by giving different sets of numerical values to the variables $x_1, x_2, \ldots, x_n$ on both sides of the equation, then we obtain an expression for $f(x_1, x_2, \ldots, x_n)$ in terms of the elementary symmetric polynomials.

EXAMPLE 6. $f(x_1, x_2, x_3) = -x_1^3 x_2 - x_1 x_2^3 - x_1^3 x_3 - x_1 x_3^3 - x_2^3 x_3 - x_2 x_3^3 + x_1^2 x_2 x_3 + x_1 x_2^2 x_3 + x_1 x_2 x_3^2$.

Here $f(x_1, x_2, x_3)$ is already a homogeneous symmetric polynomial, its highest term is $x_1^3 x_2$. We form the sets of exponents:

$$\begin{matrix} 3 & 1 & 0 \\ 2 & 2 & 0 \\ 2 & 1 & 1 \end{matrix}$$

We set $f(x_1, x_2, x_3) = -\sigma_1^2 \sigma_2 + a\sigma_2^2 + b\sigma_1\sigma_3$. When $x_1 = 1, x_2 = -1, x_3 = 0$ we have: $f(x_1, x_2, x_3) = 2, \sigma_1 = 0, \sigma_2 = -1, \sigma_3 = 0$, whence $2 = a$. When $x_1 = x_2 = x_3 = 1$ we have $f(x_1, x_2, x_3) = -3, \sigma_1 = 3, \sigma_2 = 3, \sigma_3 = 1$, whence $-3 = -27 + 18 + 3b$, $b = 2$.

$$f(x_1, x_2, x_3) = -\sigma_1^2 \sigma_2 + 2\sigma_2^2 + 2\sigma_1\sigma_3.$$

EXAMPLE 7.

$$f(x_1, x_2, \ldots, x_n) = S(x_1^2 x_2) + 2x_1^2 + 2x_2^2 + \ldots + 2x_n^2.$$

We separate $f(x_1, x_2, \ldots, x_n)$ into the homogeneous parts $f_1(x_1, x_2, \ldots, x_n) = S(x_1^2 x_2)$ and $f_2(x_1, x_2, \ldots, x_n) = 2x_1^2 + 2x_2^2 + \ldots + 2x_n^2 = S(2x_1^2)$.

In order to express f_1 in terms of the elementary symmetric polynomials, we write out the sets of exponents

$$\begin{matrix} 2 & 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 1 & 0 & \ldots & 0 \end{matrix}$$

This gives $f_1(x_1, x_2, \ldots, x_n) = \sigma_1 \sigma_2 + a\sigma_3$. Setting $x_1 = x_2 = x_3 = 1$, $x_4 = \ldots = x_n = 0$, we find $6 = 3 \times 3 + a \times 1$, $a = -3$, $f_1(x_1, x_2, \ldots, x_n) = \sigma_1 \sigma_2 - 3\sigma_3$. Similarly, we find that $f_2(x_1, x_2, \ldots, x_n) = 2\sigma_1^2 - 4\sigma_2$. Thus, $f(x_1, x_2, \ldots, x_n) = \sigma_1 \sigma_2 - 3\sigma_3 + 2\sigma_1^2 - 4\sigma_2$.

By using the fundamental theorem on symmetric polynomials, we can find the value of any symmetric polynomial in n variables for values of the variables occurring in it equal to the roots of the polynomial $g(x)$ of the nth degree in one variable, without knowing these roots themselves.

EXAMPLE 8. To find the sum of the cubes of the roots $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ of the polynomial $g(x) = 7x^4 - 14x^3 - 7x + 2$.

---

† The coefficient of the product $\sigma_1^{k_1 - k_2} \ldots \sigma_{n-1}^{k_{n-1} - k_n} \sigma_n^{k_n}$ is taken at once equal to a_0.

We take the symmetric polynomial $f(x_1, x_2, x_3, x_4) = x_1^3 + x_2^3 + x_3^3 + x_4^3$ and express it in terms of the elementary symmetric polynomials: $f(x_1, x_2, x_3, x_4) = \sigma_1^3 - 3\sigma_1\sigma_2 + 3\sigma_3$. For the values of the variables $x_1 = \alpha_1$, $x_2 = \alpha_2$, $x_3 = \alpha_3$, $x_4 = \alpha_4$, we have: $\sigma_1 = 2$, $\sigma_2 = 0$, $\sigma_3 = 1$ (see the beginning of this section). Hence $f(\alpha_1, \ldots, \alpha_4) = \alpha_1^3 + \alpha_2^3 + \alpha_3^3 + \alpha_4^3 = 8 - 0 + 3 = 11$.

EXAMPLE 9. To find the discriminant (section 1, subsection 12) of the polynomial $g(x) = x^3 - 7x - 6$.

We take the symmetric polynomial

$$f(x_1, x_2, x_3) = (x_1 - x_2)^2 (x_1 - x_3)^2 (x_2 - x_3)^2.$$

We find its highest term as the product of the highest terms of the factors, whose product is equal to $f(x_1, x_2, x_3)$ (subsection 1); we obtain $x_1^4 x_2^2$.

We construct the sets of exponents:

4	2	0
4	1	1
3	3	0
3	2	1
3	2	2

and we obtain $f(x_1, x_2, x_3) = \sigma_1^2\sigma_2^2 - a\sigma_1^3\sigma_3 + b\sigma_2^3 + c\sigma_1\sigma_2\sigma_3 + d\sigma_3^2$. Giving the variables the values $x_1 = x_2 = 1$, $x_3 = 0$, then $x_1 = x_2 = 1$, $x_3 = -2$, then $x_1 = x_2 = 2$, $x_3 = -1$, then $x_1 = x_2 = x_3 = 1$, we find in succession $b = -4$, $d = -27$, $a = -4$, $c = 18$. For values of the variables equal to the roots $\alpha_1, \alpha_2, \alpha_3$ of the polynomial $g(x)$ we obtain $f(\alpha_1, \alpha_2, \alpha_3) = D(g) = 0 + 0 + (-4)(-343) + 0 + (-27) \times 36 = 400$.

3. Expressions for power sums in terms of the elementary symmetric polynomials.

The *power sums* are the symmetric polynomials $s_k = x_1^k + x_2^k + \ldots + x_n^k$ $(k = 1, 2, \ldots)$. They are connected with the elementary symmetric polynomials (see preceding section) by *Newton's formulae* $s_k - s_{k-1}\sigma_1 + s_{k-2}\sigma_2 - \ldots + (-1)^{k-1} s_1\sigma_{k-1} + (-1)^k k\sigma_k = 0$ $(k \leqslant n)$, $s_k - s_{k-1}\sigma_1 + s_{k-2}\sigma_2 - \ldots + (-1)^n s_{k-n}\sigma_n = 0$ (when $k > n$). From these formulae we can find in turn expressions for $s_1, s_2, s_3, \ldots$, in terms of $\sigma_1, \sigma_2, \ldots, \sigma_n$ (or conversely).

EXAMPLE 1. Let $n \geqslant 3$. Then $s_1 = \sigma_1$; $s_2 - s_1\sigma_1 + 2\sigma_2 = 0$, whence $s_2 = \sigma_1^2 - 2\sigma_2$; $s_3 - s_2\sigma_1 + s_1\sigma_2 - 3\sigma_3 = 0$, whence $s_3 = \sigma_1^3 - 3\sigma_1\sigma_2 + 3\sigma_3$.

The general formulae expressing $s_k (k = 1, 2, 3, \ldots)$ in terms of $\sigma_1, \sigma_2, \ldots, \sigma_n$ ([23], p. 211) are;

$$s = k \sum (-1)^{2\lambda_1 + 3\lambda_2 + \ldots + (n+1)\lambda_n} \frac{(\lambda_1 + \ldots + \lambda_n - 1)!\, \sigma_1^{\lambda_1} \ldots \sigma_n^{\lambda_n}}{\lambda_1! \ldots \lambda!},$$

where the summation extends over all the sets of non-negative integers $\lambda_1, \lambda_2, \ldots, \lambda_n$ with the property $\lambda_1 + 2\lambda_2 + \ldots + n\lambda_n = k$ (*Waring's first formula*).

EXAMPLE 2. To find s_4.

When $k = 4$ we must take the sets (1) $\lambda_1 = 4, \lambda_2 = \ldots = \lambda_n = 0$; (2) $\lambda_1 = 2, \lambda_2 = 1, \lambda_3 = \ldots = \lambda_n = 0$; (3) $\lambda_1 = \lambda_3 = 1, \lambda_2 = \lambda_4 = \ldots = \lambda_n = 0$; (4) $\lambda_2 = 2, \lambda_1 = \lambda_3 = \ldots = \lambda_n = 0$; (5) $\lambda_1 = \lambda_2 = \lambda_3 = 0, \lambda_4 = 1, \lambda_5 = \ldots = \lambda_n = 0$. This gives $s_4 = 4$

$$\left[(-1)^8 \frac{3!\sigma_1^4}{4!} + (-1)^7 \frac{2!\sigma_1^2\sigma_2}{2!\,1!} + (-1)^6 \frac{1!\sigma_1\sigma_3}{1!\,1!} + (-1)^6 \frac{1!\sigma_2^2}{2!} + (-1)^5 \frac{\sigma_4}{1!}\right]$$
$$= \sigma_1^4 - 4\sigma_1^2\sigma_2 + 4\sigma_1\sigma_3 + 2\sigma_2^2 - 4\sigma_4.$$

Conversely, the elementary symmetric polynomials in n variables can be expressed in terms of powers sums by means of the formulae ([23], p. 213)

$$\sigma_k = \sum \frac{(1-)^{\lambda_1+\ldots+\lambda_k+k}}{1^{\lambda_1}2^{\lambda_2}\ldots k^{\lambda_k}\lambda_1!\ldots\lambda_k!}\, s_1^{\lambda_1}\ldots s_k^{\lambda_k},$$

where the summation extends over the same sets of numbers λ as in Waring's first formula (*Waring's second formula*).

EXAMPLE 3.

$$\sigma_3 = \frac{(-1)^{3+3}}{1^3\times 3!}\, s_1^3 + \frac{(-1)^{1+1+3}}{1\times 2\times 1!\,1!}\, s_1 s_2 + \frac{(-1)^{1+3}}{3\times 1!}\, s_3 = \frac{1}{6}\, s_1^3 - \frac{1}{2}\, s_1 s_2 + \frac{1}{3}\, s_3.$$

4. Systems of algebraic equations in several unknowns.

We have the system of equations

$$\left.\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0, \\ \cdot\ \cdot\ \cdot

Another special case is a system of two equations of arbitrary degree in two unknowns

$$\left.\begin{aligned} f(x, y) &= 0, \\ g(x, y) &= 0. \end{aligned}\right\} \tag{3.38}$$

To solve the system (3.38) we must associate one of the variables, for instance y, with the coefficients in each of the polynomials $f(x, y)$, $g(x, y)$, i.e. we must write these polynomials in the form

$$f(x, y) = a_0(y)x^n + a_1(y)x^{n-1} + \ldots + a_{n-1}(y)x + a_n(y),$$
$$g(x, y) = b_0(y)x^s + b_1(y)x^{s-1} + \ldots + b_{s-1}(y)x + b_s(y).$$

Then, regarding f, g as polynomials in the variable x, alone. we must construct the resultant $R(f, g)$ of these polynomials (section 1, subsection 12) and find its roots ($R(f, g)$ is a polynomial in the variable y). After this for every root β_0 of the polynomial $R(f, g)$ we must find all the common roots of the polynomials

$$\left.\begin{aligned} f(x, \beta_0) &= a_0(\beta_0)x^n + a_1(\beta_0)x^{n-1} + \ldots \\ &\quad \ldots + a_{n-1}(\beta_0)x + a_n(\beta_0), \\ g(x, \beta_0) &= b_0(\beta_0)x^s + b_1(\beta_0)x^{s-1} + \ldots \\ &\quad \ldots + b_{s-1}(\beta_0)x + b_s(\beta_0). \end{aligned}\right\} \tag{3.39}$$

If the common solutions are the numbers $\gamma_0^{(1)}, \gamma_0^{(2)}, \ldots, \gamma_0^{(k)}$, then the pairs of numbers $(\gamma_0^{(1)}, \beta_0)$, $(\gamma_0^{(2)}, \beta_0)$, $\ldots$, $(\gamma_0^{(k)}, \beta_0)$ are solutions of the system (3.38). The system (3.38) has no solutions except those obtained in the way described.

If, for the root β_0 of the resultant $R(f, g)$, $a_0(\beta_0) \neq 0$ or $b_0(\beta_0) \neq 0$, then the polynomials (3.39) are known to have at least one common root, if however $a_0(\beta_0) = 0$ and $b_0(\beta_0) = 0$, then the polynomials (3.39) cannot have common roots; in this last case no solution of the system (3.38) corresponds to the root β_0 of the resultant.

EXAMPLE 1. To solve the system

$$\left\{\begin{aligned} x^2y + x^2 + 2xy + y^3 &= 0, \\ x^2 - 6x - 3y^2 &= 0. \end{aligned}\right.$$

We associate the variable y with the coefficients and we form the resultant of the polynomials $f(x, y) = (y + 1)x^2 + 2yx + y^3$ and $g(x, y) = x^2 - 6x - 3y^2$:

$$R(f, g) = \begin{vmatrix} y+1 & 2y & y^3 & 0 \\ 0 & y+1 & 2y & y^3 \\ 1 & -6 & -3y^2 & 0 \\ 0 & 1 & -6 & -3y^2 \end{vmatrix} = y^4(4y + 3)^2.$$

The roots of the resultant are $\beta_1 = 0, \beta_2 = -\frac{3}{4}$. We consider the polynomials $f(x, 0) = x^2$ and $g(x, 0) = x^2 - 6x$; their common root is $\gamma_1 = 0$. Now we consider the polynomials $f\left(x, -\frac{3}{4}\right) = \frac{1}{4}x^2 - \frac{3}{2}x - \frac{27}{64}$ and $g\left(x, -\frac{3}{4}\right) = x^2 - 6x - \frac{27}{16}$; their common roots are $\gamma_2^{(1)} = 3 + \frac{3}{4}\sqrt{19}, \gamma_2^{(2)} = 3 - \frac{3}{4}\sqrt{19}$. The solutions of the given system of equations are: (1) $x = 0, y = 0$; (2) $x = 3 + \frac{3}{4}\sqrt{19}, y = -\frac{3}{4}$; (3) $x = 3 - \frac{3}{4}\sqrt{19}, y = -\frac{3}{4}$.

EXAMPLE 2. To solve the system

$$\begin{cases} x^2y^2 + xy^2 + x + y = 0, \\ x^2y + xy + 1 = 0. \end{cases}$$

Here it is more convenient to associate the variable x with the coefficients. We obtain polynomials $f(x, y) = (x^2 + y)y^2 + y + x$, $g(x, y) = (x^2 + y)y + 1$ and we form their resultant:

$$R(f, g) = \begin{vmatrix} x^2 + x & 1 & x \\ x^2 + x & 1 & 0 \\ 0 & x^2 + x & 1 \end{vmatrix} = (x + 1)^2x^3.$$

The roots of the resultant are $\gamma_1 = 0$, $\gamma_2 = -1$. For these values of the variable x the leading coefficients of the polynomials $f(x, y)$ and $g(x, y)$ vanish, and the polynomials obtained have no common roots ($g(0, y) = g(-1, y) \equiv 1$). Hence, the given system of equations has no solution.

For more on the solution of systems of algebraic equations in several unknowns see [13], Chapter III, section 5.

CHAPTER IV

GENERAL ALGEBRA

1. Groups

1. Definition of group. Examples.

Let G be some set ([34], pp. 23–24) of elements (for instance, a set of numbers, or functions, or any objects of a geometrical nature, and so on). We say that in the set G there is defined an *algebraic operation* if with every two elements a, b of G, taken in a definite order, some unique third element c of G can be associated. Examples of algebraic operations are: the addition of integers (to every two integers there corresponds their sum); the subtraction of integers; the addition of vectors on a plane; the addition or multiplication of square matrices of the nth order (Chapter II, section 1, subsection 1); the vector multiplication of vectors of three-dimensional space.

An algebraic operation defined in a set G is usually called either *multiplication* or *addition*. In the first case, if the element $c \in G$ corresponds to the elements a, $b \in G$, we write $c = ab$, in the second case we write $c = a + b$.

A set G with an algebraic operation defined in it (suppose that it is now called multiplication) is a *group* if the following conditions are satisfied:

(1) for any three elements a, b, $c \in G$

$$a(bc) = (ab)c$$

(the multiplication is *associative*);

(2) in the set G there exists an element e such that for any $a \in G$

$$ae = ea = a;$$

(3) for any $a \in G$ there exists an $a' \in G$ such that

$$aa' = a'a = e.$$

The element e is called the *identity* of the group, the element a' is called the element *inverse* to a, and it is usually denoted by a^{-1}. If the operation defined in the group is called addition, then instead of an identity of the group we speak of the *zero* of the group (this element is denoted by 0 and has the property $a + 0 = 0 + a = a$ for any $a \in G$), and instead of the element a^{-1}, the inverse of the element a, we speak of the element $-a$, the *additive inverse* of a $(a + (-a) = 0)$.

Examples of groups are:

(1) the set of all integers with respect to the operation of addition (the so-called *additive group* of all integers);

(2) the set of all even numbers with respect to the operation of addition (the additive group of all even numbers);

(3) the set of all non-zero rational numbers with respect to the operation of multiplication (the *multiplicative group* of non-zero rational numbers);

(4) the set of all vectors in a plane with respect to the ordinary addition of vectors;

(5) the set of all non-singular square matrices of order n with real elements with respect to the operation of multiplying matrices (Chapter II, section 1, subsection 1).

The integers with respect to the operation of multiplication do not form a group, because for an integer distinct from ± 1 there is no integer which is its inverse. For the same reason all the square matrices of order n do not form a group with respect to multiplication. The set of all vectors in three-dimensional space is not a group with respect to the vector multiplication of vectors, because this operation is not associative.

Another definition of a group (equivalent to that given above). A set G with an algebraic operation defined in it is called a group if:

(1) this operation is associative,

(2) for any two elements a, $b \in G$ there exists precisely one element $x \in G$ such that $ax = b$, and precisely one element $y \in G$ such that $ya = b$.

If the operation defined in the group is *commutative* (i.e. for any elements a, b of the group, $ab = ba$), then the group itself is called *commutative* or *Abelian*. The groups in the examples 1, 2, 3 and 4 above are Abelian, the group in Example 5 is not Abelian. Another example of a non-Abelian group is the set of all rotations of three-dimensional space about the origin of coordinates, where the product of two rotations means their successive application.

If the group G consists of a finite number of elements, then it is called a *finite group*, and the number of elements in it is called the *order of the group*.

EXAMPLE 1. The product of two permutations of the nth degree (Chapter I, section 1, subsection 1) is the result of applying them in turn (for instance, if $n = 4$ and $a = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}$, $b = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix}$, then $ab = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 3 & 1 \end{pmatrix}$, because, for instance, under the permutation a the number 1 is carried into 2, and the number 2 is carried into 4 by the permutation b, i.e. as a result 1 is carried into 4, and so on). This is an algebraic operation with respect to which the set of all permutations of the nth degree is a group. The identity of the group is the identity permutation $\begin{pmatrix} 1 & 2 & \dots & n \\ 1 & 2 & \dots & n \end{pmatrix}$, the element inverse to the permutation $\begin{pmatrix} 1 & 2 & \dots & n \\ i_1 & i_2 & \dots & i_n \end{pmatrix}$ is the permutation $\begin{pmatrix} i_1 & i_2 & \dots & i_n \\ 1 & 2 & \dots & n \end{pmatrix}$. This group is called the *symmetric group of degree* n and is usually denoted by S_n. It is a finite group of order $n!$, and when $n \geqslant 3$ it is non-commutative. (For instance, when $n = 4$ for the permutations a, b given above we have $ba = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix} \neq ab$).

If a group is finite then multiplication in it is often given by means of the *Cayley table*, in the first row and first column of which we write out all the elements of the group, beginning with the identity (let these elements be $a_1 = e, a_2, \dots, a_n$), and then at the intersection of the row in which the element a_i stands and the column in which a_j stands, we write the element equal to the product $a_i a_j$.

EXAMPLE 2. In the symmetric group of the third degree let

$$a_1 = e = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad a_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix},$$

$$a_4 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad a_5 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \quad a_6 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

Then the Cayley table has the form

	e	a_2	a_3	a_4	a_5	a_6
e	e	a_2	a_3	a_4	a_5	a_6
a_2	a_2	e	a_4	a_3	a_6	a_5
a_3	a_3	a_5	e	a_6	a_2	a_4
a_4	a_4	a_6	a_2	a_5	e	a_3
a_5	a_5	a_3	a_6	e	a_4	a_2
a_6	a_6	a_4	a_5	a_2	a_3	e

2. Isomorphism of groups.

We say that we can establish a *one-one correspondence* between the elements of the sets (see Introduction) M and N, if we can relate to every element of the set M some completely determined element of the set N, where distinct elements of N correspond to distinct elements of M and every element of N corresponds to some element of M.

Two group G and H are called *isomorphic* if we can establish a one-one correspondence between their elements, under which for any elements $a, b \in G$ and the elements $a', b' \in H$ corresponding to them, the element $c' = a' b'$ will correspond to the element $ab = c$.

EXAMPLE 1. The additive group G of all integers (subsection 1) is isomorphic to the additive group H of all even numbers (to establish an isomorphism between them we can relate the number $2k \in H$ to each number $k \in G$).

EXAMPLE 2. The multiplicative group (subsection 1) of all positive real numbers is isomorphic to the additive group of all real numbers (the isomorphism is $a \to \log a$).

An isomorphic mapping of a group onto itself is called an *automorphism* of this group.

EXAMPLE 3. One of the automorphisms of the additive group of all integers is the mapping under which we relate to every integer a the number $-a$.

If two groups are isomorphic, then the isomorphic correspondence between them can be established, generally speaking, in many ways. Isomorphic groups can be distinguished from each other only by the nature of their elements and, possibly, by the name of the operation defined in the group. But all the properties of isomorphic groups following from the properties of the operations defined in them and not depending on the nature of the elements of the group, are identical. For instance if a group is Abelian (subsection 1), then all groups isomorphic to it are also Abelian. In algebra, groups which are isomorphic to each other are not regarded as being essentially different.

3. Homomorphism.

Suppose that to every element of the group G there corresponds a uniquely defined element of the group H, where if the elements $a', b' \in H$ correspond to the elements $a, b \in G$, then the element $c' = a' b'$ corresponds to the element $ab = c$.

Such a mapping $G \to H$ is called a *homomorphism*; we say that *the group G is mapped homomorphically* into the group H. If here at least one element of the group G is mapped onto every (given) element of the group H, then we speak of a *homomorphic mapping* of the group G *onto* the group H; in this case a homomorphism is called an *epimorphism*.

In general different elements of the group G, or even no element at all, may be mapped by a homomorphism onto a given element of the group H.

EXAMPLE 1. If we make the number 1 correspond to every even number, and the number -1 to every odd number, then we obtain a homomorphic mapping of the additive group of all integers into the multiplicative group of all non-zero rational numbers (subsection 1, Examples 1 and 3). This mapping will also be a homomorphism of the additive group of all integers onto the multiplicative group consisting of the numbers -1, 1.

EXAMPLE 2. If to every non-singular square matrix of order n with real elements we relate the determinant of this matrix (Chapter I, section 1, subsection 2), then we obtain a homomorphism of the group (with respect to multiplication) of all real non-singular square matrices of order n onto the multiplicative group of all non-zero real numbers.

Under every homomorphic mapping of the group G into the group H the identity of the group G (or the zero, if the group operation is addition) goes into the identity (the zero) of the group H. The set of all elements of the group G which are carried into the identity (the zero) of the group H is called the *kernel* of the given homomorphism. In Example 1 given above the kernel of the homomorphism consists of all even numbers; in Example 2 of all matrices with determinant equal to 1.

A homomorphic mapping of a group into itself is called an *endomorphism* of this group. An example of an endomorphism is the mapping under which we relate the identity of the group to each of its elements.

If the group G is mapped homomorphically onto some set M, in which an algebraic operation is defined (subsection 1), then with respect to this algebraic operation the set M is itself necessarily a group.

4. Subgroups. Cyclic groups.

A subset A (see Introduction) of the group G is called a *subgroup* of this group if together with every element a it also contains its inverse a^{-1} and together with every two elements a, b it also contains their product ab. These two

requirements can be replaced by one: for any elements $a, b \in A$ the element ab^{-1} must lie in A.

Every subgroup of a given group G is itself a group with respect to the operation which is defined in G.

EXAMPLE 1. The additive group of all even numbers (subsection 1) is a subgroup of the additive group of all integers, which is itself a subgroup of the additive group of all complex numbers.

EXAMPLE 2. The set of all even permutations of n elements (Chapter I, section 1, subsection 1) is a subgroup of the group of all permutations of n elements (subsection 1). This subgroup is called *the alternating group of degree n* and is usually denoted by A_n.

In any group the subset consisting simply of the identity element is a subgroup. This subgroup is called the *identity subgroup* of the given group and is usually denoted by the symbol E. A group is always a subgroup of itself. Every subgroup distinct from the whole group is called a *proper subgroup* of this group.

The set of all finite groups is, essentially, the set of all subgroups of the symmetric groups (subsection 1): every finite group of order n (subsection 1) is isomorphic (subsection 2) to some subgroup of S_n *(Cayley's theorem)*.

If we take some element g in the group G and all the powers of this element† (or all its multiples if the group operation is addition), then we also obtain a subgroup of the group G. This subgroup is called *the cyclic subgroup generated by the element g* and is denoted by $\{g\}$. If the subgroup $\{g\}$ coincides with the whole group G, then the group G itself is called a *cyclic group*.

If for the element g of the group G there exists a natural number k such that $g^k = e$ (e is the identity of the group G), then g is called an *element of finite order*. The smallest natural number n with the property $g^n = e$ in this case is called *the order of the element g*. If the equation $g^k = e$ is not satisfied for any natural number k, then g is called an *element of infinite order*; in this case all the powers of the element g (including the negative ones) are different; in particular, none of these powers is equal to e, except $g^0 = e$. If the element g of the group G has infinite order, then the cyclic subgroup $\{g\}$ is infinite. If g is an element of order n, then the subgroup $\{g\}$ also has order n (subsection 1); it consists of the elements

† g^n when $n > 0$ is the product of n elements equal to g, and g^{-n} is equal to the product of n elements equal to g^{-1}, or, which is the same thing, the element inverse to g^n: $g^{-n} = (g^{-1})^n = (g^n)^{-1}$; $g^0 = e$ (by definition).

(all different) $e, g, g^2, \ldots, g^{n-1}$; every power (positive or negative) of g is equal to one of these.

EXAMPLE 3. The additive group of all integers (subsection 1) is an infinite cyclic group (it consists of all the multiples of the number 1).

EXAMPLE 4. The nth roots of 1 form (with respect to multiplication) a cyclic group of order n, generated by any of the primitive nth roots of 1 ([11], p. 128).

All infinite cyclic groups are isomorphic to each other (subsection 2). Also all finite cyclic groups of the same order n are isomorphic to each other. Every subgroup of a cyclic group is itself a cyclic group.

5. Systems of generators. Increasing sequences of subgroups.

Let M be some non-empty subset of the group G. The set of all elements of the group G which are products of (positive and negative) powers of elements of M (in each product there are a finite number of factors) is a subgroup of the group G. This subgroup is called the *subgroup generated by* the set M, and is denoted by $\{M\}$. The general form of an element of the subgroup $\{M\}$ is: $a_{\alpha_1}^{k_1} \ldots a_{\alpha_n}^{k_n}$, where every $k_i (i = 1, \ldots, n)$ is some integer, and $a_{\alpha_1}, \ldots, a_{\alpha_n}$ are elements of the set M, some of which may be the same (for instance, $a_1^2 a_2^{-1} a_1^{-3} a_3 a_2^2 \in \{M\}$, if $a_1, a_2, a_3 \in M$). The subgroup $\{M\}$ is contained in any subgroup of the group G containing the whole of the set M, and is the intersection of all subgroups of the group G containing M. The set M is called a *system of generators* of the subgroup $\{M\}$. If the subgroup $\{M\}$ coincides with the whole group G, then M is a system of generators of the group G. If here the set M is finite, then G is said to be a group with a *finite number of generators*.

EXAMPLE 1. The set of all elements of the group G is always a system of generators for it.

EXAMPLE 2. One of the systems of generators of the multiplicative group (subsection 1) of all positive rational numbers is the set of all prime numbers.

EXAMPLE 3. The number 1 is a system of generators of the additive group of all integers (subsection 1). Another system of generators of the same group consists of the numbers 2, 3.

If the set M consists of one element, then the subgroup generated by it is the cyclic subgroup generated by this element (subsection 4).

If the set M consists of all the elements belonging to several subgroups $A, B, \ldots$ of the group G, then the subgroup $\{M\}$ is called *the subgroup generated by the given subgroups* and is denoted by $\{A, B, \ldots\}$. The subgroup generated by some set of subgroups of the group G consists of all elements of the group equal to the products of elements taken from these subgroups. It is the minimal subgroup of the group G containing all the given subgroups.

Let the subgroups $A_1, A_2, \ldots, A_n, \ldots$ of the group G be such that

$$A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq \ldots \tag{4.1}$$

(*an increasing sequence of subgroups*). Then the subgroup generated by these subgroups coincides simply with the set of all elements of the group G each of which belongs to at least one of these subgroups, and is called the *union of the increasing sequence of subgroups* (4.1). We can assume that the union of an increasing sequence of subgroups of the group G coincides with this group itself.

EXAMPLE 4. The additive group (subsection 1) of all rational numbers is the union of the increasing sequence of cyclic subgroups

$$\{1\} \subset \left\{\frac{1}{2}\right\} \subset \left\{\frac{1}{6}\right\} \subset \ldots \subset \left\{\frac{1}{n!}\right\} \subset \ldots$$

EXAMPLE 5. The set of all complex numbers which are roots of 1 ([11], p. 127) of degree equal to a power of some fixed prime number p form a group with respect to multiplication. This group is the union of the increasing sequence of (cyclic) subgroups $A_1 \subset A_2 \subset \ldots$, where $A_i (i = 1, 2, \ldots)$ consists of all the p^ith roots of 1. This group (and every group which is isomorphic to it) is called a *group of type* p^∞ or a *quasi-cyclic* group. The groups of type p^∞ play an important part in the theory of Abelian groups.

6. Cosets. Decomposition of a group with respect to a subgroup.

If A is a subgroup, and g an arbitrary element of the group G, then by gA we denote the set of all elements of the group G obtained by multiplying the element g by all possible elements of the subgroup A (i.e. the set of all elements of the form ga, where $a \in A$). This set is called the *left coset of the group G with respect to the subgroup A determined by the element g*. Similarly the *right coset* Ag of the group G with respect to the subgroup A determined by the element g is the set of all elements of the form ag, where $a \in A$. If the group operation is addition, then the left and right cosets are denoted respectively by $g + A$ and $A + g$.

EXAMPLE 1. In the group of all permutations of degree 3 (subsection 1) we take the subgroup A consisting of the permutation $e = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$ and $a = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$, and consider the element $g = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$. Then the left coset gA will consist of the permutations $ge = g = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ and $ga = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$, and the right coset Ag will consist of the permutations $eg = g = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ and $ag = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$.

Every left coset is determined by any of the elements contained in it, i.e. if $g_1 \in gA$, then $g_1 A = gA$. Any two left cosets of the group G with respect to the subgroup S either coincide or else have no elements in common. One of the left cosets of the group G with respect to the subgroup A is the subgroup A itself ($A = eA$, where e is the identity of the group G); all the remaining left cosets with respect to the subgroup A are not subgroups of the group G. Similar assertions are true for right cosets.

The number of different left cosets of the group G with respect to the subgroup A is always equal to the number of different right cosets of the group G with respect to this subgroup (in the infinite case this means that the cardinal number of the set of all the different left cosets of the group G with respect to the subgroup A coincides with the cardinal number of the set of right cosets). This number (in the infinite case—cardinal number) is called the *index* of the subgroup A in G. If G is finite, then its order (subsection 1) is equal to the product of the order of any of its subgroups A by the index of this subgroup in G (*Lagrange's theorem*). From this it follows that the order of every subgroup of a finite group is a divisor of the order of the group. Also the order of any element (subsection 4) of a finite group is a divisor of the order of this group. Conversely, if the order of a finite group G is divisible by the prime number p, then G has elements of order p (*Cauchy's theorem*).

If we choose a subgroup A in an arbitrary group G, then the elements of G can be decomposed into non-intersecting classes, by combining those elements which lie in the same left coset of the group G with respect to the subgroup A. Such a decomposition is called the left *decomposition of the group G with respect to the subgroup A*. If instead of the left

cosets we take the right cosets with respect to the subgroup A, then we obtain the *right decomposition of the group G with respect to the subgroup A.*

EXAMPLE 2. In the symmetric group of degree 3 S_3 (subsection 1), the elements of which are $a_1 = e = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$, $a_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$, $a_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$, $a_4 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$, $a_5 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$, $a_6 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$, we take the subgroup A consisting of the elements e and a_6. Then in the left decomposition of S_3 with respect to A the set of its elements is decomposed into the classes: (1) e, a_6 (the coset $eA = A$), (2) a_2, a_5 (the coset $a_2 A$), (3) a_3, a_4 (the coset $a_3 A$), and in the right decomposition we obtain the classes: (1) e, a_6 (the coset $Ae = A$), (2) a_2, a_4 (the coset Aa_2), (3) a_3, a_5 (the coset Aa_3). The index of A in S_3 is 3.

EXAMPLE 3. Both in the left and in the right decompositions of the additive group of all integers (subsection 1) with respect to the subgroup consisting of all integers divisible by 3, each coset will consist of those integers which when divided by 3 have the same remainder. In all we obtain three classes: (1) the class B_0 of numbers divisible by 3 without remainder, (2) the class B_1 of numbers which have remainder 1 when divided by 3, and (3) the class B_2 of numbers which have remainder 2 when divided by 3.

If in an arbitrary group G we take G itself as the subgroup A, then (both the right and the left) decompositions of the group G with respect to the subgroup A will consist of just one coset (which contains all the elements of G). If, however, we take the identity subgroup E as the subgroup A (subsection 4), then every element of G will form a separate coset.

7. Normal subgroups of a group.

If in the left and right decompositions of the group G with respect to some subgroup H of it (subsection 6) the classes into which the elements of G are divided turn out to be identical, then H is called a *normal subgroup* of G (or an *invariant subgroup*).

EXAMPLE 1. In Example 2 of subsection 6 the subgroup A of S_3 is not a normal subgroup.

EXAMPLE 2. If the index of the subgroup H in the group G is 2, then H is a normal subgroup of G: in the left and in the right decomposition of G with respect to H one of the cosets is H itself, and the other coset consists of all the elements of G not lying in H. For instance, the alternating group of degree n (subsection 4), whose index in the symmetric group S_n (subsection 1) is 2, is a normal subgroup.

A subgroup H is a normal subgroup of the group G if and only if for any element g of G

$$gH = Hg.$$

This equation means that for every element h of H we can find elements h' and in h'' H such that

$$gh = h'g, \quad hg = gh''. \tag{4.2}$$

EXAMPLE 3. If the group G is Abelian (subsection 1), then every subgroup H of it is a normal subgroup (to obtain the equation (4.2) it is sufficient to take $h' = h'' = h$).

The elements g_1 and g_2 of the group G are called *conjugate* in this group if in G there exists an element g such that

$$g_2 = g^{-1}g_1g$$

(in this case we also say that the element g_2 is obtained from g_1 by *transforming* by the element g).

If A is a subgroup, and g a fixed element of some group G, then the set of all elements of the form $g^{-1}ag$, where a varies over the whole of A, is again a subgroup of G. This subgroup is called a subgroup *conjugate* to A in G.

EXAMPLE 4. The subgroup A_1 of S_3 in subsection 6, Example 2, consisting of the elements e and a_2 is conjugate to the subgroup A considered there: $A_1 = a_3^{-1}Aa_3$.

A subgroup H of the group G is a normal subgroup of this group if and only if together with every element h it contains all the elements conjugate to h in G.

EXAMPLE 5. Let G be the group of all non-singular square matrices of order n with real elements (subsection 1, Example 5), H the subgroup consisting of all matrices whose determinants (Chapter I, section 1, subsection 2) are equal to 1. Then if the matrix A lies in H, every matrix of the form $\mathbf{B}^{-1}\mathbf{AB}$ ($\mathbf{B}$ is any non-singular matrix) also lies in H because $|\mathbf{B}^{-1}\mathbf{AB}| = |\mathbf{B}^{-1}||\mathbf{A}||\mathbf{B}| = |\mathbf{A}| = 1$ (pp. 21–56). Hence, H is a normal subgroup of the group G.

The normal subgroups of the group G, and only they, are the subgroups conjugate to themselves in G. Every normal subgroup of a group G is the kernel of some homomorphism (subsection 3) of this group; conversely, the kernel of every homomorphism of G is a normal subgroup of G.

In any group the identity subgroup (subsection 4) and the group itself are normal subgroups. If there are no other normal subgroups, then it is called a *simple* group. An example of

a simple group is the alternating group of the degree n (subsection 4) when $n \geqslant 5$. There also exist infinite simple groups.

If all the subgroups in the group G are normal subgroups, and the group G is not Abelian (subsection 1), then it is called a *Hamiltonian* group. For an example of such a group see [12], p. 58.

8. Factor-group.

If in the set of all cosets of the group G with respect to the normal subgroup H (subsections 6 and 7) we introduce an operation by means of the rule $(g_1 H)(g_2 H) = (g_1 g_2)H$ (in the additive notation $(g_1 + H) + (g_2 + H) = (g_1 + g_2) + H$), then this will be an algebraic operation† (subsection 1), under which this set of cosets itself becomes a group. This group is called the *factor-group of G with respect to the normal* subgroup H and is denoted by G/H. The identity of the factor-group G/H is the coset H. The element inverse to an element g^H of the factor-group is the coset $g^{-1}H$. The order of the factor-group G/H is equal to the index of H in G (subsections 1 and 6).

EXAMPLE 1. The factor-group of the group of all integers (subsection 1) with respect to the subgroup of all numbers divisible by 3 consists of three elements—the classes B_0, B_1 and B_2 (subsection 6, Example 3). This is a cyclic group of order 3, generated by the element B_1 (subsection 4) (the group operation is addition).

EXAMPLE 2. In the multiplicative group G of all non-zero rational numbers (subsection 1) the numbers 1, -1 form a normal subgroup H. The cosets are the pairs of numbers c, $-c$, where c is some positive rational number. The product of the cosets c_1, $-c_1$ and c_2, $-c_2$ is the coset of G with respect to H, consisting of the numbers $c_1 c_2$, $-c_1 c_2$ ($c_1 H c_2 H = (c_1 c_2)H$). The factor-group G/H is isomorphic (subsection 2) to the multiplicative group of all positive rational numbers (this isomorphism can be obtained by setting up a correspondence between the number c and the coset c, $-c$).

EXAMPLE 3. If in the multiplicative group G of all non-singular square matrices of the order n with real elements we take the normal subgroup H consisting of all the matrices whose determinants are equal to 1 (subsection 7, Example 5), then each coset with respect to H will consist of all matrices with given determinant. If to every coset we relate the determinant of all the matrices in this coset, then we obtain an isomorphism between the factor-group H/G and the multiplicative group of all non-zero real numbers (subsection 1 and 2).

---

† If $g_1' H = g_1 H$, $g_2' H = g_2 H$, then $(g_1' g_2' H = (g_1 g_2)H$, i.e. the coset $(g_1 g_2)H$ does not depend on the choice of the elements g_1, g_2 in the given cosets $g_1 H$, $g_2 H$.

EXAMPLE 4. Let G be the additive group of all vectors in the plane, starting from the origin of coordinates, H its subgroup consisting of all the vectors lying along the ordinate axis. Then H is a normal subgroup of G, and every coset of G with respect to H is the set of all vectors on the plane, starting from the origin of coordinates and having their endpoints on some (fixed) straight line parallel to the ordinate axis. The factor-group G/H is isomorphic (subsection 2) to the additive group of all real numbers (to establish the isomorphism between these groups we can set up a correspondence between every coset of G with respect to H and the real number which is represented by the point of intersection of the straight line on which the endpoints of all the vectors of the given coset lie, with the abscissa axis).

If to every element g of the group G we relate the coset gH with respect to some normal subgroup H, then we obtain a homomorphic mapping of the group G onto the factor-group G/H. It is called the *natural homomorphic mapping of G onto G/H*. The kernel of this homomorphism (subsection 3) is H. The set of all elements of G which are carried by the natural homomorphism into the same fixed element of G/H is a coset of G with respect to H.

If the group G can be mapped homomorphically onto some group $\overline{G}$ and H is the kernel of this homomorphism, then H is necessarily a normal subgroup of G, and the group $\overline{G}$ is isomorphic (subsection 2) to the factor-group G/H (*the theorem on the homomorphism of groups*).

EXAMPLE 5. Let $\{a\}$ be an infinite cyclic group, and $\{b\}$ a cyclic group of order n (subsection 4). If to every element a^k of the group $\{a\}$ (k is any integer) we relate the element b^k of the group $\{b\}$, then we obtain a homomorphic mapping of the group $\{a\}$ onto the group $\{b\}$. The kernel of this homomorphism is the cyclic subgroup of $\{a\}$ generated by the element $\{a^n\}$. The factor-group $\{a\}/\{a^n\}$ is isomorphic to the group $\{b\}$.

9. Groups of permutations ([27]; [23]; [5], Chapter I).

Groups of permutations are subgroups of the symmetric groups (subsection 1). In studying them it is usually convenient to use expressions for the permutations in the form of products of cycles. A *cycle* or *cyclic permutation* is a permutation of the numbers $1, 2, \ldots, n$† which carries one of these numbers i_1 into the number i_2, i_2 into i_3 and so on, i_{k-1} into i_k ($k \leqslant n$),

† Instead of a permutation of the numbers $1, 2, \ldots, n$ one can speak of permutations of any n elements $\alpha_1, \ldots, \alpha_n$; if instead of these elements one writes only the indices by which they are numbered, then the result is exactly the same as if we were discussing permutations of numbers.

i_k into i_1, and all the remaining numbers are left unchanged. This permutation is denoted by $(i_1, i_2, \ldots, i_k)$. The cycles $(i_1, i_2, \ldots, i_k)$ and, for instance, $(i_2, \ldots, i_k, i_1)$ are equal to each other. The number k is called the *length* of the cycle. A cycle of length 1 is the identity permutation. In order to represent an arbitrary permutation of the numbers $1, 2, \ldots, n$ in the form of a product of cycles, we take any of these numbers (for instance, 1), then the number into which it is carried by the given permutation (let this be the number j_2), then the number into which j_2 is carried by this permutation, and so on up to the number j_l which is carried by the given permutation into the first of the numbers chosen (in our case, into 1), and we write the cycle $(1, j_2, \ldots, j_l)$; after this we must take any of the numbers $1, 2, \ldots, n$ which has not yet occurred (if such a number exists), and, beginning with it, do the same, and so on until we have used all the numbers $1, 2, \ldots, n$. Then we obtain several cycles, (which have no numbers in common which they actually permute) the product of which (in any order) will be equal to the given permutation. Cycles which do not contain any numbers in common which they really permute are called *independent*. Cycles of length 1 occurring in a product are usually omitted. Under this condition, every permutation can be decomposed into the product of independent cycles in a unique way.

EXAMPLE 1.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 8 & 2 & 1 & 5 & 7 & 4 & 6 & 3 \end{pmatrix} = (183)(4576).$$

EXAMPLE 2. The permutations forming the symmetric group of degree 3 can be written in the form of products of cycles: $a_1 = (1)$, $a_2 = (2, 3)$, $a_3 = (1,2)$, $a_4 = (1, 2, 3)$, $a_5 = (1, 3, 2)$, $a_6 = (1, 3)$.

A cycle of length 2 is called a *transposition*. Every permutation can be represented in the form of a product of transpositions (for instance, $(i_1, i_2, \ldots, i_k) = (i_1, i_2,)(i_1, i_3) \ldots (i_1, i_k)$). This representation is not unique, but every even permutation (Chapter I, section 1, subsection 1) can be decomposed into the product of an even number only of transpositions, and every odd permutation can be decomposed into the product of an odd number only. Every even permutation can be represented in the form of a product of cycles of length 3 (for instance, $(1, 2)(1, 3) = (1, 2, 3)$; $(1, 2)(3, 4) = (1, 3, 2)(2, 4, 3)$).

The order (subsection 4) of a cycle as an element of the group of permutations in which it lies is always equal to its length. The order of a permutation, represented in the form

of a product of independent cycles, is the least common multiple of the lengths of these cycles.

If the permutation a is represented in the form of a product of cycles, then, in order to obtain a permutation $b^{-1}ab$ conjugate to it, we apply the permutation b to the numbers forming these cycles.

EXAMPLE 3.

$$a = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 6 & 4 & 5 & 3 & 2 \end{pmatrix} = (2, 6)(3, 4, 5);$$

$$b = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 3 & 1 & 6 & 5 & 2 \end{pmatrix}; \quad b^{-1}ab = (3, 2)(1, 6, 5).$$

A group of permutations of n elements $\alpha_1, \ldots, \alpha_n$ is called *transitive* if for any two of these elements α_i, α_j the group contains a permutation carrying α_i into α_j. If this condition is not satisfied, the group is called *intransitive*.

EXAMPLE 4. The symmetric group of degree n (subsection 1) is transitive. The alternating group of degree n (subsection 4) is also transitive if $n \geqslant 3$.

EXAMPLE 5. The subgroup of the symmetric group on n symbols ($n > 2$) consisting of the permutations (1), (1, 2) is intransitive.

A transitive group of permutations of the elements $\alpha_1, \ldots, \alpha_n$ is called *imprimitive* if the set of elements $\alpha_1, \ldots, \alpha_n$ can be separated into non-intersecting subsets (*systems of imprimitivity*) in such a way that each of these subsets contains more than one element and every permutation of the group carries any one of these subsets either into itself, or into one of the other subsets.

All the systems of imprimitivity consist of the same number of elements. If the set $\alpha_1, \ldots, \alpha_n$ cannot be separated into systems of imprimitivity, then the group is called *primitive*.

EXAMPLE 6. The symmetric group of any degree and the alternating group of degree at least 3 are primitive.

EXAMPLE 7. If the number n is prime, then every transitive group permutation on n elements is primitive.

EXAMPLE 8. The subgroup of the group S_4, consisting of the permutations (1), (1, 2) (3, 4), (1, 3) (2, 4), (1, 4) (2, 3) (the so-called *Klein four-group*), is imprimitive; systems of imprimitivity are, for instance, the subsets 1, 2 and 3, 4.

In any group P of permutations of the elements $\alpha_1, \ldots, \alpha_n$ the set P_{α_i} of all permutations which leave a fixed element α_n unchanged is a subgroup. If the group P is transitive, then

the subgroups $P_{\alpha_1}, \ldots, P_{\alpha_n}$ are conjugate to each other in P (subsection 7), and the index of each of them in P (subsection 6) is n. From this it follows that the order (subsection 1) of a transitive group of permutations of n elements is always divisible by n (subsection 6).

A transitive group P of permutations on the elements $\alpha_1, \ldots, \alpha_n$ is primitive if and only if the subgroups P_{α_i} $(i = 1, \ldots, n)$ are maximal proper subgroups. If this condition is not satisfied so that there exists a subgroup Q of the group P such that $P_{\alpha_i} \subset Q \subset P$ for some i, then systems of imprimitivity for the group P are the subsets of the set $\alpha_1, \ldots, \alpha_n$, each of which consists of all the elements into which α_i (i fixed) is carried by the permutations belonging to the same right coset of the group P with respect to the subgroup Q (subsection 6).

10. Direct products of groups.

The group G is called the *direct product* (or the *direct sum*, if the group operation is addition) of its subgroups $H_1, \ldots, H_n$, if:

(1) every subgroup $H_i (i = 1, \ldots, n)$ is a normal subgroup (subsection 7) of G;

(2) G is generated (subsection 5) by the subgroups $H_1, \ldots, H_n$;

(3) the intersection (see Introduction) of each subgroup $H_i (i = 1, \ldots, n)$ with the subgroup generated by all the $H_j, j \neq i$, contains only the identity of the group.

Requirement (3) can be replaced by a weaker one: the intersection of each subgroup $H_i (i = 2, \ldots, n)$ with the subgroup generated by $H_1, \ldots, H_{i-1}$ must contain only the identity.

Another definition of a direct product (equivalent to the first). The group G is called the direct product of its subgroups $H_1, \ldots, H_n$, if:

(1) every element g of the group G can be expressed uniquely in the form of a product

$$g = h_1 \ldots h_n,$$

where $h_i \in H_i$ $(i = 1, \ldots, n)$;

(2) the elements of any two subgroups H_i and H_j $(i \neq j)$ commute with each other (i.e. $h_i h_j = h_j h_i$, if $h_i \in H_i$, $h_j \in H_j$).

If the group G is the direct product (direct sum) of its subgroups $H_1, H_2, \ldots, H_n$, then we write: $G = H_1 \times H_2 \times \ldots$

$\times H_n$ (or $G = H_1 \dot{+} H_2 \dot{+} \ldots \dot{+} H_n$). The subgroups $H_1, H_2, \ldots, H_n$ are called the *direct factors* (*direct summands*) occurring in the given direct decomposition of the group G.

If $G = H_1 \times \ldots \times H_n$ and $G \ni g = h_1 \ldots h_n$ ($h_i \in H_i$, $i = 1, \ldots, n$), then the elements $h_1, \ldots, h_n$ are called the *components* of the element g with respect to the direct factors $H_1, \ldots, H_n$. For any i ($i = 1, \ldots, n$) the component of the product of two elements of the group G in H_i is equal to the product of the components of these elements in H_i, i.e. if $g = h_1 \ldots h_n$, $g' = h'_1 \ldots h'_n$, then $gg' = (h_1 h'_1) \ldots (h_n h'_n)$.

EXAMPLE 1. The Klein four-group V (subsection 9) is the direct product of its subgroup H_1, consisting of the permutations (1) and (1, 2) (3, 4), and its subgroup H_2, consisting of the permutations (1) and (1, 3) (2, 4). It is also the direct product of H_1 and the subgroup H_3, consisting of the permutations (1) and (1, 4) (2, 3), and also the direct product of the subgroups H_2 and H_3: $V = H_1 \times H_2 = H_1 \times H_3 = H_2 \times H_3$.

EXAMPLE 2. The real square matrices of order n (see Introduction) with positive determinants (Chapter I, section 1, subsection 2) form a group G with respect to the multiplication of matrices (Chapter II, section 1, subsection 1) in which the set H_1 of all matrices with determinants equal to 1 is a normal subgroup. The set H_2 of all scalar matrices (Chapter II, section 1, subsection 3) with positive numbers in the diagonal is also a normal subgroup of G. In the intersection of H_1 and H_2 ([34], pp. 23–4) there is only the identity matrix, which is the identity of the group G; $G = H_1 \times H_2$.

If $G = H_1 \times H_2 \times \ldots \times H_n$, where every H_i ($i = 1, \ldots, n$) is itself the direct product $H_i = H_{i1} \times \ldots \times H_{ik_i}$, then we can obtain a new direct decomposition for G: $G = H_{11} \times \ldots \times H_{1k_1} \times \ldots \times H_{n1} \times \ldots \times H_{nk_n}$ (a *continuation* of the first direct decomposition of G). Another direct decomposition of G can be obtained by separating, in an arbitrary way, the set of subgroups $H_1, \ldots, H_n$ which occur in the original direct decomposition, into non-intersecting parts and replacing the subgroups contained in each of these parts by their product (i.e. by the subgroup generated by them (subsection 5)).

EXAMPLE 3. Let G be the group (with respect to addition) of all vectors in three-dimensional space, starting from the origin of coordinates, H_1, H_2, H_3 its subgroups consisting, respectively, of all vectors along the x-axis, the y-axis and the z-axis, starting from the origin of coordinates, A and B the subgroups consisting of the vectors on the plane xOy and the vectors on the plane yOz. Then $A = H_1 \dot{+} H_2$, $B = H_2 \dot{+} H_3$, $G = A \dot{+} H_3 = H_1 \dot{+} H_2 \dot{+} H_3 = H_1 \dot{+} B$.

If $G = H_1 \times H_2$, then the factor-group G/H_1, (subsection 8) is isomorphic (subsection 2) to the subgroup H_2; similarly,

the factor-group G/H_2 is isomorphic to H_1. If C is a subgroup of G which contains the direct factor H_1, then $C = H_1 \times D$, where D is the intersection ([34], pp. 23–4) of the subgroups C and H_2. If $G = H_1 \times \ldots \times H_n$ and H_i' is a subgroup of H_i $(i = 1, \ldots, n)$, then the subgroup generated in G by all the subgroups H_i' (subsection 5) is their direct product $\{H_1', \ldots, H_n'\} = H_1' \times \ldots \times H_n'$.

A definition of the direct product (direct sum) can also be given in the case of an *infinite number of direct factors* (direct summands): the group G is called *the direct product (direct sum) of some set of its subgroups* H_α (notation $G = \prod_\alpha H_\alpha$ or, respectively, $G = \dot{\sum}_\alpha H_\alpha$), if G is generated by these subgroups and if the subgroup generated in G by any finite number of the subgroups H_α is their direct product (direct sum). The properties of direct products mentioned above are also true for the case of an infinite number of factors.

EXAMPLE 4. The multiplicative group of all non-zero rational numbers (subsection 1) is the direct product of the subgroup consisting of the numbers -1, 1 and an infinite number of infinite cyclic subgroups, each of which is generated by some prime number.

EXAMPLE 5. The group (with respect to addition) of all polynomials (of all degrees) in one variable x with real coefficients is the direct sum of its subgroups, each of which consists of all polynomials of the form ax^k where k is fixed ($k = 0, 1, 2, \ldots$), and the coefficient a varies over all the real numbers.

If $H_1, \ldots, H_n$ are any groups, then we can always construct a group which is the direct product of subgroups isomorphic (subsection 2) to $H_1, \ldots, H_n$. Such a group is obtained if we take the set of all sequences of the form $(h_1, \ldots, h_n)$, where h_i is an element of the group H_i $(i = 1, \ldots, n)$, and we multiply the sequences together term by term: $(h_1, \ldots, h_n) \times (h_1', \ldots, h_n') = (h_1 h_1', \ldots, h_n h_n')$ $(h_i, h_i' \in H_i, i = 1, \ldots, n)$.

If we are given an infinite set of groups H_α, then in the same way we can construct a group which is the direct product of groups isomorphic to the H_α. Its elements will be systems of elements h_α, taken one from each of the groups H_α, where in each of these systems all the elements except a finite number must be the identities of the corresponding groups.

11. Abelian groups.

When we are considering Abelian groups (subsection 1), the group operation is usually called addition. An Abelian group G is the direct sum (subsection 10) of its subgroups

H_α (α varies over some set of indices) if and only if each of its elements g can be expressed uniquely in the form of a sum of a finite number of elements taken from different groups H_α.

EXAMPLE 1. The additive group (subsection 1) of all complex numbers is the direct sum of the additive group of all real numbers and the additive group of all purely imaginary numbers ([11], p. 115).

Every Abelian group with a finite number of generators (subsection 5) is the direct sum of a finite number of infinite cyclic (subsection 4) subgroups and a finite number of cyclic subgroups whose orders (subsection 4) are powers of prime numbers (*the fundamental theorem for finitely generated Abelian groups*).

EXAMPLE 2. The cyclic group of order 12 generated by the element a is the direct sum of the cyclic subgroup of order 4, generated by the element $3a$, and the cyclic subgroup of order 3, generated by the element $4a$: $\{a\} = \{3a\} \dotplus \{4a\}$.

An Abelian group which cannot be decomposed into the direct sum of proper subgroups is called *indecomposable*. Infinite cyclic groups and cyclic groups whose orders are powers of prime numbers are indecomposable.

If an Abelian group with a finite number of generators can be decomposed in two ways into the direct sum of infinite cyclic groups and cyclic groups whose orders are powers of prime numbers, then in both decompositions the number of infinite cyclic summands, the number of finite cyclic summands and the set of their orders will be the same.

Every subgroup A of an Abelian group G with a finite number of generators is itself a group with a finite number of generators. If here in the decomposition of the group G into the direct sum of infinite cyclic groups and cyclic groups of prime-power orders there are r infinite cyclic summands ($r \geqslant 0$) and to the prime number p there correspond summands of orders $p^{\alpha_1}, \ldots, p^{\alpha_{k_p}}$, where

$$\alpha_1 \geqslant \ldots \geqslant \alpha_{k_p},$$

(for all k_p of the summands, $k_p \geqslant 0$), and in the decomposition of the group A into the direct sum of infinite cyclic groups and cyclic groups of prime-power orders, there are s infinite cyclic summands and to the prime number p there correspond the summands of orders $p^{\beta_1}, \ldots, p^{\beta_{l_p}}$ ($\beta_1 \geqslant \ldots \geqslant \beta_{l_p}$), then necessarily $s \leqslant r$ and for any prime number p, $l_p \leqslant k_p$ and $\beta_1 \leqslant \alpha_1, \ldots$

$\ldots, \beta_{l_p} \leqslant \alpha_{l_p}$. On Abelian groups, not necessarily having a finite system of generators, see [12], Chapters VI–VIII.

2. Rings

1. Definitions. Examples (see [5]).

A set R with two algebraic operations defined in it (section 1, subsection 1), addition and multiplication, is called a *ring* if with respect to the operation of addition it is an Abelian group (section 1, subsection 1), and the operation of multiplication is related to the operation of addition by the *distributive* laws, i.e. for any three elements $a, b, c \in R$

$$a(b+c) = ab + ac \quad \text{and} \quad (b+c)a = ba + ca. \tag{4.3}$$

Multiplication defined in a ring is not necessarily either associative or commutative. If the multiplication defined in R is associative (i.e. for any three elements $a, b, c \in R$ $a(bc) = (ab)c$), then the ring R is called an *associative ring*. If, in addition, the multiplication defined in R is commutative, then R is called a *commutative ring*. In a commutative ring the second of the equations (4.3) is a consequence of the first.

If in the ring R the condition $a^2 = 0$ is satisfied for any element a and for any three elements a, b, c

$$a(bc) + b(ca) + c(ab) = 0$$

(*Jacobi's identity*), then R is called a *Lie ring*.

EXAMPLE 1. The integers form a commutative ring with respect to the usual operations of addition and multiplication of numbers.

EXAMPLE 2. The rational numbers, the real numbers, the complex numbers form commutative rings with respect to the usual operations of addition and multiplication.

EXAMPLE 3. All polynomials in one variable with arbitrary numerical coefficients form a commutative ring with respect to the usual operations of addition and multiplication of polynomials.

EXAMPLE 4. An associative non-commutative ring is formed by all square matrices of order n with arbitrary numerical elements (if again the addition and multiplication of matrices are performed in the usual way, see Chapter II, section 1, subsection 1).

EXAMPLE 5. The set of all vectors of three-dimensional space, where vectors are added in the usual way, and the product of two vectors is their vector product, is a non-associative ring. This ring is a Lie ring.

The Abelian group, which is obtained if we consider the operation of addition in a ring only, is called the *additive group* of the ring. The zero element of this group is called the *zero of the ring*. The product of any element a of the ring by zero is equal to zero: $a.0 = 0$. If for the elements a, b of some ring, $ab = 0$, but $a \neq 0$ and $b \neq 0$, then a and b are called *divisors of zero* (a is a left divisor, b a right divisor of zero). If there are no divisors of zero in R, then R is called a *ring without divisors of zero*. A commutative ring without divisors of zero is called an *integral domain*.

EXAMPLE 6. Every ring in which the elements are numbers, and the operations are the ordinary addition and multiplication of numbers, is an integral domain.

EXAMPLE 7. The set of all functions defined and continuous ([34], p. 48) in the segment $[-1, 1]$ is a ring with divisors of zero with respect to the ordinary operations of addition and multiplication of functions. (For instance, the product of the functions

$$f_1(x) = \begin{cases} 0 \text{ when } -1 \leqslant x \leqslant 0, \\ x \text{ when } \quad 0 < x \leqslant 1; \end{cases} \qquad f_2(x) = \begin{cases} x \text{ when } -1 \leqslant x \leqslant 0, \\ 0 \text{ when } \quad 0 < x \leqslant 1, \end{cases}$$

neither of which is equal to the zero of the ring, is zero.)

If for the elements a, b_1, b_2 of the ring the equation $ab_1 = ab_2$ (or $b_1 a = b_2 a$) is satisfied, where $a \neq 0$ is not a left (respectively, a right) divisor of zero, then $b_1 = b_2$, i.e. equations can be cancelled by non-zero elements which are not divisors of zero. It is impossible to cancel by an element which is a divisor of zero.

EXAMPLE 8. In the ring of all square matrices of order 2 (Example 4) for the matrices

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 1 & 2 \\ 4 & 0 \end{pmatrix}$$

the equation $AB_1 = AB_2$ is true, although $B_1 \neq B_2$. Here A is a left divisor of zero: for instance, $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

An element e of the ring R is called the *identity* of this ring if for any element $a \in R$, $ae = ea = a$. It may also be that the ring does not have an identity. If there is an identity in the ring R, then R is called a *ring with identity*.

EXAMPLE 9. The ring of all integers is a ring with identity.

EXAMPLE 10. The even numbers form a ring without identity.

In a ring with identity e there may exist for the element $a \neq 0$ an *inverse* element a^{-1} with the property $aa^{-1} = a^{-1}a = e$ (but it may also be that there is no such element). Elements of a ring with identity for which there exist inverse elements in this ring are called *units*.

EXAMPLE 11. In the ring of all square matrices of order n the identity is the identity matrix (Chapter II, section 1, subsection 2). An inverse element exists for every non-singular matrix (Chapter II, section 1, subsection 2); for singular matrices there do not exist any inverse matrices.

2. Isomorphism. Homomorphism.

The rings R and Q are called *isomorphic* if we can establish a one–one correspondence between their elements (section 1, subsection 2) such that if the elements $b_1, b_2 \in Q$ correspond to the elements $a_1, a_2 \in R$, then the element $b_1 + b_2$ corresponds to $a_1 + a_2$ and the element $b_1 b_2$ corresponds to the elements $a_1 a_2$.

EXAMPLE 1. The ring of all square matrices of order n with real elements (subsection 1, Example 4) is isomorphic to the ring of all linear transformations operating in real n-dimensional linear vector space (addition and multiplication of transformations is defined in the usual way: see [8], [11], [14], [16], [32]).

We say that the ring R is *mapped homomorphically* into the ring Q if a uniquely defined element of the ring Q can be made to correspond to each element of the ring R, such that if the elements $b_1, b_2 \in Q$ correspond to the elements a_1, a_2 of the ring R, then the element $b_1 + b_2$ corresponds *to* $a_1 + a_2$, the element $b_1 b_2$ to the element $a_1 a_2$. If here every element of the ring Q is the image of some element of R, then we speak of a homomorphic mapping of the ring R onto the ring Q; the homomorphism in this case is called an *epimorphism*.

EXAMPLE 2. If to every polynomial in one variable with any complex coefficients we relate the free term of this polynomial, then we obtain a homomorphic mapping of the ring of all polynomials with complex coefficients onto the ring (in fact the field) of all complex numbers.

If the ring R is homomorphically (in particular, isomorphically) mapped onto the ring Q, and R is associative (subsection 1), then Q is also associative; if R is commutative, then Q is also commutative; if R is a ring with identity (subsection 1), then Q is also a ring with identity, and moreover the identity of Q is the image of the identity of R. If R can be homomorphically (or isomorphically) mapped onto some set M with two

algebraic operations (section 1, subsection 1), then M itself is a ring with respect to these two operations.

If the ring R is mapped homomorphically into the ring Q, then the set of all elements of R which are mapped into the zero of Q is called the *kernel* of this homomorphism.

3. Subrings. Ideals.

A subgroup A (section 1, subsection 4) of the additive group of R (subsection 1) is called a *subring* of this ring if together with every two of its elements a_1, a_2 it also contains their product $a_1 a_2$. A subring A of R is called a *left* (respectively, a *right*) *ideal* of this ring if together with every element a it also contains all elements of the form ra (of the form ar), where r is an arbitrary element of R. In commutative rings the concepts of left and right ideals coincide. If a subring A of an arbitrary ring is both a left and a right ideal, then it is called a *two-sided ideal.*

EXAMPLE 1. In the ring of integers, the multiples of some fixed number n form a two-sided ideal.

EXAMPLE 2. In the ring of all square matrices of order n the set of all matrices in which the last column consists of zeros is a left ideal, and the set of all matrices in which the last row consists of zeros is a right ideal.

In any ring the element O and the ring itself are two-sided ideals. If there are no other two-sided ideals in the ring, it is called a *simple ring.* An example of a simple ring is the ring of all square matrices of order n with any complex (or with any real) elements (subsection 1, Example 4). Every field (section 3) is also a simple ring.

If in the ring R there is given some set (see Introduction) of elements N, then the smallest (left, right or two-sided) ideal of R containing all the elements of N is called the (left, right or two-sided) *ideal generated by the set* N; this ideal is the intersection ([34], pp. 23–24) of all the (left, right or two-sided) ideals of R containing N. An ideal generated by one element a is called a *principal ideal* (notation: $(a)_l$, $(a)_r$ or (a), depending on whether the ideal is left, right or two-sided). If R is a commutative ring, then the principal ideal (a) generated by the element a of this ring consists of all elements of the form $ra + na$, where $r \in R$, n is an integer. If R is a commutative ring with identity, then (a) consists simply of all elements of the form ra, where $r \in R$.

If R is a ring with identity e, then it is itself the principal ideal generated by the element e. An integral domain (subsection 1) with identity in which every ideal is principal is called a *principal ideal ring*. For instance, the following are principal ideal rings: the ring of integers, the ring of polynomials in one variable with coefficients from any field (Chapter III, section 1).

If R is a principal ideal ring and ε is a unit of R (subsection 1), then any element $a \in R$ can be represented in the form $a = a\varepsilon^{-1}\varepsilon$. Such a decomposition of the element a is called *trivial*. If for the element $p \neq 0$ of R there are no non-trivial decompositions (i.e. from the equation $p = cd$ it always follows that c or d is a unit), then p is called an *indecomposable* or *prime element* (for instance, the prime numbers in the ring of integers, a polynomial irreducible over a given field (Chapter III, section 3, subsection 7) in the ring of all polynomials with coefficients from this field).

In a principal ideal ring every non-zero element is the product of prime elements. If here $a = p_1 p_2 \dots p_s = q_1 q_2 \dots q_t$ are two decompositions of the element a into prime factors (which are not units), then $s = t$ and the p_i coincide with the q_j to within the order of numbering and to within unit factors.

4. Direct sums of rings.

If the additive group of an arbitrary ring R (subsection 1) can be decomposed into the direct sum of its subgroups A_α (section 1, subsection 10), where α varies over some set of indices, and every A_α is a two-sided ideal (subsection 3) of the ring R, then we say that the ring R can be *decomposed into a* (two-sided) *direct sum*. If we are given some set of rings R_α, then we can always construct a ring which is the two-sided direct sum of ideals isomorphic to the R_α. For this we must construct the Abelian group which is the direct sum of the subgroups $\overline{R}_\alpha$ isomorphic to the additive groups of the rings R_α (section 1, subsection 10) and define multiplication in it by the formula

$$(a_{\alpha_1} + \dots + a_{\alpha_n})(a'_{\alpha_1} + \dots + a'_{\alpha_n}) = a_{\alpha_1} a'_{\alpha_1} + \dots + a_{\alpha_n} a'_{\alpha_n}$$

($a_{\alpha_i}, a'_{\alpha_i} \in \overline{R}_{\alpha_i}$, $i = 1, \dots, n$, where some of the elements a_{α_i} or a'_{α_i} may also be zero); we find the product $a_{\alpha_i} a'_{\alpha_i}$ by using the rule for multiplication in the ring R_{α_i}.

5. Factor-ring.

Let R be an arbitrary ring, A a two-sided ideal of it. Then we can construct the factor-group (section 1, subsection 8) of the additive group of R (subsection 1) with respect to the subgroup consisting of all elements of A. If in the set of all cosets $r + A$ $(r \in R)$, which are the elements of this factor group, we introduce another operation, (multiplication) by means of the rule $(r + A)(r' + A) = rr' + A$ $(r, r' \in R)$, then we obtain a ring, called the *factor-ring of R with respect to A* (notation: R/A).

EXAMPLE. The factor-ring of the ring of integers with respect to the ideal A consisting of all even numbers contains two elements: the coset $a_0 = 0 + A = A$ and the coset $a_1 = 1 + A$. These elements are added by the rule $a_0 + a_0 = a_0$, $a_0 + a_1 = a_1 + a_0 = a_1$, $a_1 + a_1 = a_0$ and are multiplied by the rule $a_0 a_0 = a_0$, $a_0 a_1 = a_1 a_0 = a_0$, $a_1 a_1 = a_1$.

If to every element r of R we relate the coset $r + A$ with respect to the two-sided ideal A determined by it, then this gives a homomorphic mapping (subsection 2) of R onto the factor-ring R/A, the kernel of which will be the ideal A (subsection 2). If the ring R can be mapped homomorphically on some ring Q and A is the kernel of this homomorphism, then A is necessarily a two-sided ideal of R, and the factor-ring R/A is isomorphic to Q (*the homomorphism theorem for rings*).

3. Fields. Skew Fields

1. Fields (see [5]).

A *field* is a commutative ring (section 2, subsection 1), not consisting of zero alone, in which for any element $a \neq 0$ and any element b there exists precisely one element x such that $ax = b$. The element x is called the *quotient* after the division of the element b by the element a (notation: $x = b/a$).

Examples of fields are the ring of rational numbers, the ring of real numbers, the ring of complex numbers. All complex numbers which are roots of polynomials with rational coefficients also form a field, called the *field of algebraic numbers*. The ring of integers is not a field. But every factor-ring of it (section 2, subsection 5) with respect to an ideal consisting of all numbers which are multiples of some prime number p is a field (*the field of residue classes modulo p*). Another example of a field is that of all rational functions $f(x)/g(x)$, where $f(x)$ and $g(x)$ are polynomials with real coefficients, and $g(x) \neq 0$.

Every field has an identity (section 2, subsection 1), and for any non-zero element of a field there exsists an element inverse to it (section 2, subsection 1). The set of all non-zero elements of a field form an Abelian group (section 1, subsection 1) with respect to the multiplication defined in the field (*the multiplicative group of the field*). No field contains divisors of zero (section 2, sul section 1). The only ideals of a field (section 2, subsection 3) are the zero ideal and the field itself.

A set with two algebraic operations (section 1, subsection 1) isomorphic to a field (section 2, subsection 2) is itself a field. Every homomorphic mapping of one field into another (section 2, subsection 2) is either an isomorphism, or a mapping which carries all the elements of the field into zero.

If some positive integral multiple of the identity element e of the field $Pne = e + e + \ldots + e$ (n summands) is equal to zero, then the least positive integer p with the property $pe = 0$ is called the *characteristic of the field* P; p is always a prime number. If no positive integral multiple of the identity element of the field P is equal to zero, then P is called a field *of characteristic zero*. An example of a field of characteristic p is the field of residue classes modulo p, an example of a field of characteristic zero is any number field (for instance, the field of all real numbers).

The set P' of elements of the field P is called a *subfield* of this field if it is itself a field with respect to the operations which are defined in P; P' is a subfield of P if and only if together with any two elements a, b it also contains $a+b$, ab, $a-b$† and a/b (if $b \neq 0$). If P' is a subfield of P, then P is called an *extension* of P'. A field which has no subfields except itself is called a *simple field*. Examples of simple fields are the field of residue classes modulo p and the field of all rational numbers.

If in the field P we are given the subfield P' and the set of elements N, then the smallest subfield P'' of P containing P' and N is the field obtained by *adjoining* the set N to P' (notation: $P'' = P'(N)$). The field $P'(N)$ consists of all elements obtained from elements of P' and N by addition, subtraction, multiplication and division, and is the intersection (see Introduction) of all subfields of P containing P' and N. If N consists of one element α, then P'' is called a *simple extension* of P'. If here α is a root of some polynomial $f(x)$ with coefficients from the field P' (Chapter III, section 1), then $P'' = P'(\alpha)$

† $a - b$ is the element which is the solution of the equation $b + x = a$.

is called a *simple algebraic extension* of P', and the element α is said to be *algebraic with respect to* P'; if there is no polynomial $f(x)$ with coefficients from P' for which the element α is a root, then $P'' = P'(\alpha)$ is called a *simple transcendental extension* of the field P' and the element α is said to be *transcendental* with respect to P'. If the simple algebraic extension $P'(\alpha)$ of the field P' is obtained by adjoining the element α to it, which is a root of the polynomial $f(x)$ of degree n with coefficients from P', irreducible over P' (Chapter III, section 3, subsection 7), then every element of $P'(\alpha)$ can be written uniquely in the form $a_0 + a_1\alpha + \ldots + a_{n-2}\alpha^{n-2} + a_{n-1}\alpha^{n-1}$, where $a_0, \ldots, a_{n-1}$ are elements of P'. If $P'(\eta)$ is a simple transcendental extension of the field P', then each of its elements has the form $f(\eta)/g(\eta)$, where $f(\eta)$ and $g(\eta) \neq 0$ are polynomials in η with coefficients from P'.

EXAMPLE 1. A simple algebraic extension of the field of rational numbers obtained by adjoining to it the number $\alpha = \sqrt[3]{2}$, consists of all numbers of the form $a + b\sqrt[3]{2} + c\sqrt[3]{4}$ where a, b, c are rational numbers.

EXAMPLE 2. The field of all complex numbers is a simple algebraic extension of the field of all real numbers, obtained from it by adjoining the root i of the polynomial $x^2 + 1$.

EXAMPLE 3. The field of all rational functions of the form $f(x)/g(x)$, where $f(x)$ and $g(x)$ are polynomials with rational coefficients ($g(x) \neq 0$), is a simple transcendental extension of the field of rational numbers.

If $f(x)$ is a polynomial with coefficients from a field P (Chapter III, section 1), then there always exists an extension $\overline{P}$ of P such that $f(x)$ is the product of polynomials of the first degree with coefficients from $\overline{P}$, $f(x) = a_0(x - \alpha_1) \ldots (x - \alpha_n)$, where $a_0 \in P$, $\alpha_i \in \overline{P}$, $i = 1, \ldots, n$. The subfield of $\overline{P}$ obtained by adjoining the elements $\alpha_1, \ldots, \alpha_n$ to P is called the *decomposition field* of the polynomial $f(x)$ ($\alpha_1, \ldots, \alpha_n$ are the roots of $f(x)$). If all the roots of any polynomial with coefficients from P lie in P itself, then P is said to be *algebraically closed*. Examples of algebraically closed fields are the field of all complex numbers, the field of algebraic numbers.

If every element of some extension $\tilde{P}$ of the field P is algebraic with respect to P, then $\tilde{P}$ is called an *algebraic extension* of P. Every simple algebraic extension of P is an algebraic extension. For any field there exists an algebraic extension which is an algebraically closed field, i.e. no further algebraic extension is possible (*Steinitz's theorem*).

2. Skew fields.

A *skew field* is an associative ring (section 2, subsection 1), not consisting of zero alone, in which for any element $a \neq 0$ and any element b there exist an element x and an element y such that $ax = b$, $ya = b$. The difference between a skew field and a field lies in the fact that multiplication in a skew field is not necessarily commutative. An example of a skew field which is not a field is the skew field of quaternions (section 4).

Sometimes in the definition of a skew field the requirement of the associativity of the ring is not included.

4. Algebras

A ring A (section 2, subsection 1) is called an *algebra* over the field P (section 3), if its additive group (section 2, subsection 1) is a vector space ([34], p. 58) over P and if multiplication in A is connected with multiplication by elements of P by the formula $\alpha(ab) = (\alpha a)b = a(\alpha b)$ $(a, b \in A,\ \alpha \in P)$ ([34], p. 23).

If the vector space formed by the additive group of the algebra A is n-dimensional ([34], p. 58), then the number n is called the dimension of the algebra A. An algebra of finite dimension is also called a *hypercomplex system.* If the algebra A is a Lie ring (section 2, subsection 1), then it is called a *Lie algebra* ([19], p. 379.) If an algebra is not only a ring, but is also a skew field, then it is called a *division algebra.*

EXAMPLE 1. The set of all vectors of three-dimensional space, in which vectors are added and multiplied by numbers in the usual way, and the product of two vectors is their vector product, is a Lie algebra over the field of real numbers.

EXAMPLE 2. The set of all square matrices of order n with complex elements, in which the operations of addition and multiplication of matrices and the operation of multiplying matrices by a complex number are defined in the usual way (Chapter II, section 1, subsection 1), is an algebra of dimension n over the field of complex numbers.

EXAMPLE 3. The complex numbers form a division algebra of dimension 2 over the field of real numbers with respect to the ordinary operations of addition and multiplication of complex numbers and the ordinary operation of multiplying complex numbers by real numbers.

EXAMPLE 4. Every field is an algebra of dimension 1 over itself.

If the vector space over the field P which is the additive group of an algebra A has basis $e_1, \ldots, e_n$ ([34]. p. 56), then it is sufficient to know the products of the elements of this basis by each other in order to know the product of any element of the algebra A by any other element of it. (If $a, b \in A$, where $a = \alpha_1 e_1 + \ldots + \alpha_n e_n$, $b = \beta_1 e_1 + \ldots + \beta_n e_n$ (α_i, $\beta_i \in P, i = 1, \ldots, n$), then $ab = \alpha_1\beta_1 e_1^2 + \alpha_1\beta_2 e_1 e_2 + \ldots + \alpha_1\beta_n e_1 e_n + \alpha_2\beta_1 e_2 e_1 + \ldots + \alpha_n\beta_n e_n^2$.) Hence, we give a *multiplication table* for an algebra of finite dimension, where, for any two elements e_i, e_j of the basis of the vector space which is the additive group of the algebra, we give a linear combination of the elements of this basis, equal to the product $e_i e_j$.

EXAMPLE 5. If in the algebra of all complex numbers over the field of real numbers we take as basis the elements 1, i, then the multiplication table will look like this:

	1	i
1	1	i
i	i	-1.

EXAMPLE 6. If in the algebra of all square matrices of order n over the field of complex numbers we take as the basis the matrices $E_{ij} (i, j = 1, \ldots, n)$, where at the intersection of the ith row and the jth column there is a one, and the remaining elements are zeros, then the products in this algebra will be given by the multiplication table

$$E_{ij}E_{kl} = \begin{cases} E_{il} & \text{when } j = k, \\ 0 & \text{when } j \neq k. \end{cases}$$

If the multiplication of the elements of the basis is associative or commutative, then in general the multiplication defined in the algebra is also associative or commutative respectively.

The algebra whose additive group is the four-dimensional vector space over the field of real numbers with basis $1, i, j, k$ and in which multiplication is given by the multiplication table

	1	i	j	k
1	1	i	j	k
i	i	-1	k	$-j$
j	j	$-k$	-1	i
k	k	j	$-i$	-1,

is called the *algebra of quaternions*. This is a non-commutative, but associative division algebra. Its elements are called *quaternions*. They are usually written in the form $z = \alpha \cdot 1 + \beta i$

$+\gamma j+\delta k$, where $\alpha, \beta, \gamma, \delta$ are real numbers. The quaternion $\bar{z}=\alpha.1-\beta i-\gamma j-\delta k$ is said to be *conjugate* to z; $z\bar{z}=\alpha^2+\beta^2+\gamma^2+\delta^2=|z|^2$ (a generalization of the square of the modulus of a complex number). The inverse for the quaternion $z\neq 0$ is the quaternion $z^{-1}=\bar{z}/|z|^2$.

The field of real numbers, the field of complex numbers and the algebra of quaternions are the only associative division algebras of finite dimension over the field of real numbers (*Frobenius' theorem*) ([16], p. 368).

Two algebras A and A' over the same field P are said to be *isomorphic* if we can establish a one–one correspondence between their elements, which gives an isomorphism of the rings A and A' (section 2, subsection 2), where if the element $a'\in A'$ corresponds to the element $a\in A$, then for any $\alpha\in P$ $\alpha a'$ corresponds to αa. We say that the algebra A over the field P is mapped *homomorphically* into the algebra C over the same field if the ring A is mapped homomorphically into the ring C (section 2, subsection 2), where if the element $c\in C$ corresponds to the element $a\in A$, then for any $\alpha\in P$ the element αc corresponds to the element αa.

EXAMPLE 7. The algebra of all complex numbers over the field of real numbers is isomorphic to the algebra of all square matrices of order 2 of the form $\begin{pmatrix}\alpha & \beta\\ -\beta & \alpha\end{pmatrix}$ with real elements α, β.

A sub-ring B (section 2, subsection 3) of the algebra A over the field P is called a *sub-algebra* of this algebra if its additive group (section 2, subsection 1) is a sub-space ([34], p. 58) of the space formed by the additive group of the algebra A. If the sub-algebra B is a left, right or two-sided ideal of the ring A (section 2, subsection 3), then it is called, respectively, a *left-*, *right-* or *two-sided ideal of the algebra* A. The set B of elements of the algebra A is a left ideal of this algebra if and only if, together with every two of its elements b_1 and b_2 it contains their difference b_1-b_2 and, together with every element b, it contains the product ab of any element $a\in A$ by the element b and the product αb of any element of the field P by the element b.

EXAMPLE 8. In the algebra of quaternions the elements of the form $\alpha.1+\beta i$ form a sub-algebra (isomorphic to the algebra of all complex numbers over the field of real numbers).

EXAMPLE 9. In the algebra of all square matrices of order 2 over the field of complex numbers, the matrices of the form $\begin{pmatrix}0 & \alpha\\ 0 & \beta\end{pmatrix}$ form a left ideal.

EXAMPLE 10. In every algebra with an identity e (section 2, subsection 1) over the field P (not consisting of zero alone) the set of all elements of the form αe, where $\alpha \in P$, is a sub-algebra, isomorphic to the field P.

If B is a two-sided ideal of the algebra A over the field P, then we can take the factor-ring A/B (section 2, subsection 5). If in this factor-ring we introduce an operation of multiplication by elements from the field P by the rule $\alpha(a+B)= =\alpha a+B$ $(\alpha \in P, a \in A)$, then we obtain an algebra over the field P, called the *factor-algebra* of the algebra A with respect to the ideal B.

5. Lattices

1. Partially ordered sets.

A set S is said to be *partially ordered* if for some of its elements a, b there is defined a relation $a \leqslant b$ (in words "a is less than or equal to b", or "a precedes b", or "a is contained in b"), and the following conditions are satisfied:

(1) it is always true that $a \leqslant a$;

(2) from $a \leqslant b$ and $b \leqslant a$ it follows that $a = b$ (i.e. the elements a and b coincide);

(3) from $a \leqslant b$, $b \leqslant c$ it follows that $a \leqslant c$ (transitivity).

If $a \leqslant b$, but $a \neq b$, then we write $a < b$. Instead of $a \leqslant b$ we can also write $b \geqslant a$ and say "b is greater than or equal to a" or "b follows a", or "b contains a"; instead of $a < b$ we can also write $b > a$. If $b > a$ and in the given partially ordered set there is no element x with the property $b > x > a$. then we say that b *covers* a.

EXAMPLE 1. The set S consists of all subsets A, B, ... of some set M (including the empty set), and we decree that $A \leqslant B$ if every element of the subset A also lies in B (the empty set is regarded as being contained in any of the others).

EXAMPLE 2. The set S consists of all the subgroups $A, B, \ldots$, of some group G (section 1, subsection 4), and the relation $A \leqslant B$ means that A is contained in B.

EXAMPLE 3. The set S consists of all the natural numbers, and the relation $a \leqslant b$ means that the number a is a divisor of the number b.

EXAMPLE 4. The set S consists of all real single-valued functions ([34], p. 48), defined in the segment [0, 1], where it is assumed that $f(x) \leqslant g(x)$ if at every point of the segment [0, 1] $f(x)$ takes a value smaller than or equal to the value of $g(x)$ at this point.

A partially ordered set is said to be fully *ordered* (or *linearly ordered*), if for every pair of its elements a, b either $a \leqslant b$ or $b \leqslant a$. For instance, the set of all real numbers, where $a \leqslant b$ means that the number a is less than or equal to the number b.

If the set S is partially ordered (fully ordered) by means of some relation $\leqslant$, then any of its subsets is also partially ordered (fully ordered) by means of the same relation.

If we can establish a one–one correspondence between the elements of two partially ordered sets S and S' (section 1, subsection 2), such that for any $a, b \in S$ and the elements $a', b' \in S'$ corresponding to them the relation $a' \leqslant b'$ always follows from the relation $a \leqslant b$ and conversely, then the sets S and S' are called *isomorphic*. An isomorphism of a partially ordered set with itself is called an *automorphism*.

EXAMPLE 5. If S is the set of all real numbers and $a \leqslant b$ means that the number a is less than or equal to the number b, then the correspondence $a \to a + 2$, for instance, is an automorphism of the set S.

If in the partially ordered set S we replace the given ordering by that opposite to it, i.e. if we assume that $b \leqslant a$ if and only if in the first ordering $a \leqslant b$, then we again obtain an ordered set, called the *dual* of the set S.

EXAMPLE 6. The set S^* consisting of all the subsets $A, B, \ldots$, of some set M, where we assume that $A \leqslant B$ if every element of the subset B lies in A, is the dual of the set S of Example 1.

2. Fundamental definitions ([2]).

A partially ordered set S (subsection 1) is called a *lattice* if for any two of its elements a, b there exist elements $c, d \in S$ with the properties: (1) $c \leqslant a, c \leqslant b$; (2) if $c' \leqslant a$ and $c' \leqslant b$ $(c' \in S)$, then necessarily $c' \leqslant c$; (3) $d \geqslant a$, $d \geqslant b$; (4) if $d' \geqslant a$ and $d' \geqslant b$ $(d' \in S)$, then necessarily $d' \geqslant d$.

The element c is called the *intersection* of the elements a and b (notation: $c = a \cap b$), the element d is called their *union* (notation: $d = a \cup b$).

EXAMPLE 1. The partially ordered sets of Examples 1–4 of the preceding section are lattices, where in the first of them the intersection and union of the subsets of the set M are their intersection and union in the set-theoretical sense ([34], pp. 23–24), in the second example the intersection of two subgroups is their intersection in the set-theoretical sense, and the union is the subgroup generated (section 1, subsection 5) by the two given subgroups, in the third example the intersection is the greatest common divisor, and the union the least common multiple of the two given numbers,

in the fourth example the intersection (union) of the functions $f(x)$ and $g(x)$ is the function which takes at each point of the segment [0, 1] the least (greatest) of the values taken at this point by the functions $f(x)$ and $g(x)$.

In any lattice S for every finite set of its elements $a_1, \ldots, a_n$ there exist elements c, d such that $c \leqslant a_i \leqslant d$ $(i = 1, \ldots, n)$ and if $c' \leqslant a_i$ $(i = 1, \ldots, n)$, $c' \in S$ (or $d' \geqslant a_i$, $i = 1, \ldots, n$, $d' \in S$), then $c' \leqslant c$ (respectively, $d' \geqslant d$); c is called the *intersection*, d the *union* of the elements $a_1, \ldots, a_n$. If an intersection and a union exist for every set of elements of the lattice S (possibly infinite), then the lattice S is called *complete*. Every finite lattice is complete. Examples of complete lattices are the lattice of all subgroups of a given group (Example 1), the lattice of all subrings (section 2, subsection 3) of a given ring (where for the subrings A, B it is assumed that $A \quad B$ if A is contained in B).

If in the lattice S there is an element a_0 such that for every $b \in S$ it is true that $a_0 \leqslant b$, then a_0 is called the *zero* of S and is denoted by 0. If in S there is an element a_1 such that for any $b \in S$ it is true that $b \leqslant a_1$, then a_1 is called the *identity* of the lattice and is denoted by 1. In any lattice consisting of a finite number of elements the zero is the intersection, and the identity is the union of all the elements of the lattice. In the lattice of all subgroups of a given group the zero is the identity subgroup (section 1, subsection 4), and the identity is the whole group. In the lattice of Example 4 of the preceding section there is no zero and no identity.

A subset (see Introduction) of a lattice which contains, together with any two of its elements, their intersection and their union is called a *sublattice* of the given lattice. Clearly, every sublattice is itself a lattice with respect to the partial ordering given in the whole lattice. But a subset of a lattice may be a lattice with respect to this same partial ordering even in the case when it is not a sublattice.

EXAMPLE 2. The lattice of all normal subgroups (section 1, subsection 7) of an arbitrary group G (where for the normal subgroups H, K we assume that $H \leqslant K$ if H is contained in K) is a sublattice of the lattice of all its subgroups.

EXAMPLE 3. The lattice of all subgroups of a group (Example 1) is not a sublattice of the lattice of all subsets of the set of all the elements of this group.

Two lattices are called *isomorphic*, if they are isomorphic as partially ordered sets (subsection 1). Every partially ordered set isomorphic to a lattice is itself a lattice.

If a lattice consists of a finite number of elements, then we can construct a *diagram* for it: the elements of the lattice are represented by points on a plane in such a way that if $b > a$, then the point representing b lies above the point representing a; points representing elements of which one covers the other (subsection 1) are joined by straight lines.

EXAMPLE 4. Let the group G be the direct product (section 1, subsection 10) of two of its cyclic subgroups of order 2 (section 1, subsection 4): $G = \{a\} \times \{b\}$. Then the lattice of subgroups of the group G has the diagram illustrated in Fig. 1, where $A = \{a\}$, $B = \{b\}$, $C = \{ab\}$, E is the identity subgroup (section 1, subsection 4).

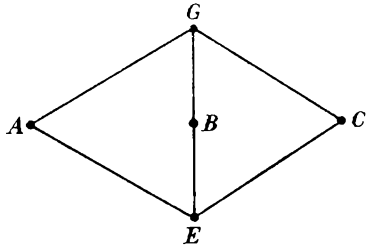


FIG. 1.

Another definition of a lattice. A set S is called a lattice if in it there are defined two algebraic operations (section 1, subsection 1), multiplication and addition, which relate to every pair of elements $a, b \in S$ their product ab and their sum $a + b$, where these operations are commutative and associative (i.e. $ab = ba$, $a+b = b+a$, $a(bc) = (ab)c$, $a + (b+c) = (a+b)+c$), and which for every $a \in S$ satisfy the conditions $aa = a$, $a+a = a$ and are connected with each other by the condition if $ab = a$, then $a + b = b$, and conversely.

This definition of a lattice is equivalent to that given above. To prove this it is necessary to regard the intersection and union of the elements a, b (in the sense of the first definition) as, respectively, their product and sum and, conversely, to assume that $a \leqslant b$ if $ab = a$ (in the sense of the second definition).

3. Distributive and modular lattices.

If in the lattice S addition and multiplication are connected by the distributive laws, i.e. for any three elements $a, b, c \in S$

$$(a + b)c = ac + bc, \tag{4.4}$$

then the lattice S is called *distributive*.

The distributive law in a lattice can also be expressed in the form $(a + c)(b + c) = ab + c$.

A lattice S is distributive if and only if in it the equation $a = b$ follows from the equations $ac = bc$, $a + c = b + c$ $(a, b, c \in S)$.

EXAMPLE 1. Every lattice which is a fully ordered set (subsection 1) is distributive (for instance, the lattice of all integers, where $a \leqslant b$ means that the number a is less than or equal to the number b). The sets of Examples 1, 3 and 4, subsection 1, are also distributive lattices. The lattice of all subgroups of a given group G is distributive if and only if G is either a cyclic group or the union of an increasing sequence of cyclic groups (section 1, subsections 4 and 5).

Every sub-lattice (subsection 2) of a distributive lattice is distributive. A lattice which is the dual (subsection 1) of a distributive lattice is also distributive.

A lattice is not distributive if and only if it contains a sub-lattice (subsection 2) which has the form of one of the two diagrams illustrated in Figs. 2 and 3.

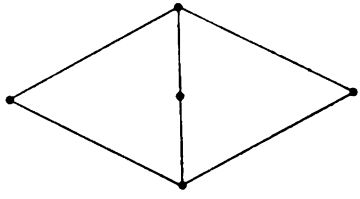


FIG. 2.

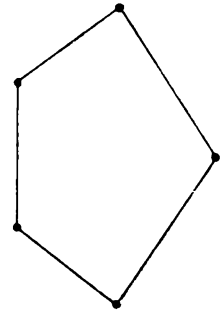


FIG. 3.

If in the lattice S the equation (4.4) is satisfied under the condition that $a \leqslant c$ (and without this condition it may still be satisfied), then the lattice S is called *modular* (or *Dedekind*). A lattice S is modular if and only if for any elements $a, b, c \in S$ the equation $(ac + b)c = ac + bc$ is satisfied. Another necessary and sufficient condition for the lattice S to be modular is: for any elements $a, b, c \in S$ the equation $a = b$ must follow from the equations $ac = bc$, $a + c = b + c$ and the relation $a \leqslant b$.

EXAMPLE 2. The lattice of all normal subgroups of an arbitrary group is modular.

EXAMPLE 3. Every lattice which has the diagram illustrated in Fig. 2 is modular.

Every sub-lattice of a modular lattice and the dual of a modular lattice are themselves modular lattices.

A lattice S is not modular if and only if it contains a sublattice having the form of the diagram illustrated in Fig. 3.

4. Boolean algebras.

A lattice S, containing a 0 and 1 (subsection 2), is called a *complemented* lattice if for each of its elements a there exists an element $b \in S$ such that $a \cap b = 0$, $a \cup b = 1$. A distributive complemented lattice (subsection 3) is called a *Boolean algebra*.

The lattice of all subsets of a given set (subsection 2, Example 1) is always a Boolean algebra. Conversely, every *finite* Boolean algebra is isomorphic (subsection 2) to the lattice of all the subsets of some set. An arbitrary Boolean algebra is isomorphic to a sublattice of the lattice of all subsets of some set.

In a Boolean algebra the complement of every element is unique; the complement of the intersection (union) of two elements is equal to the union (intersection) of the complements of each of these elements. In any Boolean algebra S the infinite distributive law is valid: if for an infinite set of elements $\ldots, a_\alpha, \ldots$ of S there exists in S the union $\bigcup_\alpha a_\alpha$ and for some element $a \in S$ there exists in S the union $\bigcup_\alpha (a \cap a_\alpha)$, then

$$a \cap \left(\bigcup_\alpha a_\alpha\right) = \bigcup_\alpha (a \cap a_\alpha).$$

A *sub-algebra* of a Boolean algebra is a sublattice of it (subsection 2), which contains together with each of its elements a, its complement a'.

If in a Boolean algebra we call the intersection of two elements their *product*, and the element $(a \cap b') \cup (a' \cap b)$ (a' and b' are the complements of the elements a and b respectively) the *sum* of the elements a, b, then we obtain a commutative ring with 1 (section 2, subsection 1), for each element a of which $2a = 0$ and $a^2 = a$ (a so-called *Boolean ring with identity*).

Conversely, if in a Boolean ring with identity we assume that $a \geqslant b$ if $b = ab$, then we obtain a Boolean algebra, where $a \cap b = ab$ (in the sense of the multiplication defined in the ring) and $a \cup b = a + b - ab$.

BIBLIOGRAPHY

1. ALEKSANDROV, P. S., *Introduction to the Theory of Groups*, Moscow, Uchpedgiz, 1938.
2. BIRKHOFF, G., *Lattice Theory*, Moscow, IL, 1952.
3. BOCHER, M., *Introduction to Higher Algebra*, Moscow, Gostekhizdat, 1933.
4. BRONSHTEIN, I. N. and SEMENDYAYEV, K. A., *A Guide-Book to Mathematics for Technologists and Engineers*, Pergamon Press, 7th edition, 1964.
5. VAN DER WAERDEN, B. L., *Modern Algebra*, Vols. 1 and 2, Gostekhizdat, 1947.
6. GANTMAKHER, F. R., *Matrix Theory*, Gostekhizdat, 1953.
7. GANTMAKHER, F. R. and KREIN, M. G., *Oscillating Matrices and Kernels and Small Vibrations of Mechanical Systems*, Gostekhizdat, 2nd edition, 1950.
8. GELFAND, I. M., *Lectures in Linear Algebra*, Gostekhizdat, 2nd edition, 1951.
9. JACOBSON, N., *The Theory of Rings*, IL, 1947.
10. KAGAN, V. F., *Fundamental Theory of Determinants*, State Publishing Company of the Ukraine, Odessa, 1922.
11. KUROSH, A. G., *A Course in Higher Algebra*, 6th edition, Fizmatgiz, 1959.
12. KUROSH, A. G., *The Theory of Groups*, 2nd edition, Gostekhizdat, 1953.
13. LYAPIN, YE. S., *A Course in Higher Algebra*, 2nd edition, Uchpedgiz, 1955.
14. MAL'TSEV, A. I., *The Elements of Linear Algebra*, 2nd edition, Gostekhizdat, 1956.
15. MEIMAN, N. N., Some problems on the position of roots of polynomials, *UMN*, **4**, 34 (1949).
16. OKUNEV, L. YA., *Higher Algebra*, 4th edition, Gostekhizdat, 1949.
17. OKUNEV, L. YA., The ring of polynomials and the field of rational functions, in: *Encyclopedia of Elementary Mathematics*, Book 2, Gostekhizdat, 1951.
18. OKUNEV, L. YA., *The Elements of Modern Algebra*, Uchpedgiz, 1941.
19. PONTRYAGIN, L. S., *Continuous Groups*, 2nd edition, Gostekhizdat, 1954.

20. PROSKURYAKOV, I. V., *Collected Problems in Linear Algebra*, Gostekhizdat, 1957.
21. SMIRNOV, V. I., *A Course in Higher Mathematics*, Vol. 1, 16th edition, Gostekhizdat, 1956. (English translation by Pergamon Press, Oxford, 1964.)
22. *Modern Mathematics for Engineers*, Edited by BEKKENBAKH, E. F., IL, 1958.
23. SUSHKEVICH, A. K., *The Elements of Higher Algebra*, 4th edition, Gostekhizdat, 1941.
24. FADDEYEV, D. K. and SOMINSKII, I. S., *Collected Problems in Higher Algebra*, 2–6 editions, Gostekhizdat, 1949–55.
25. FADDEYEV, D. K. and FADDEYEVA, V. N., *Computational Methods in Linear Algebra*, Fizmatgiz, 1960.
26. FRAZER, R., DUNCAN, V. and COLLAR, A., *The Theory of Matrices and its Applications to Differential Equations and Dynamics*, IL, 1950.
27. CHEBOTAREV, N. G., *The Elements of Galois Theory*, Part I, ONTI, 1934.
28. CHEBOTAREV, N. G., *Galois Theory*, ONTI, 1936.
29. CHEBOTAREV, N. G. and MEIMAN, N. N., *The Routh-Hurwitz Problem for Polynomials and Integral Functions*, Trudy Matem. in-ta im. V. A. Steklova, Vol. XXVI, izd. AN SSSR, 1949.
30. CEZARÒ E., *An Elementary Textbook of Algebraic Analysis and the Calculation of Infinitesimals*, Part I, ONTI, 1936.
31. SHAPIRO, G. M., *Higher Algebra*, 4th edition, Uchpedgiz, 1938.
32. SHILOV, G. YE., *Introduction to the Theory of Linear Spaces*, Gostekhizdat, 1952.
33. SCHMIDT, O. YU., *The Abstract Theory of Groups*, 2nd edition, ONTI, 1933.
34. *Mathematical Analysis (functions, limits, series, continued fractions)*, SMB, Moscow, Fizmatgiz, 1961.

INDEX OF SPECIAL SYMBOLS

$-a$ the additive inverse of the element a of a group 255

A_n the alternating group of degree n 229

$\tilde{D}$ the associated determinant 26

$\tilde{\mathbf{A}}$ the associated matrix 99

$A(\mathbf{x}; \mathbf{y})$ a bilinear form 117

$I_a^b \Phi(x)$ the Cauchy index of the rational function $\Phi(x)$ between the limits from a to b 204

$\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_s\}$ a cellular diagonal matrix 73

A_{ij} the cofactor of an element 8

$\mathbf{A}^*$ the conjugate matrix 104

$(a_1 a_2 a_3 \ldots a_n)$ a continuant 15

$(a_1, a_2, \ldots, a_k)$ a cycle (cyclic permutation) 236

$\{g\}$ the cyclic group generated by the element g 229

$|\mathbf{A}|$ the determinant of the matrix $\mathbf{A}$ 4

$\begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix}$, $|a_{ij}|_1^n$ a determinant of the nth order 4

$\{a_1, a_2, \ldots, a_n\}$ a diagonal matrix 72

$A \times B$, $\prod A_\alpha$ the direct product of groups 239

$A \dot{+} B$, $\dot{\sum} A_\alpha$ the direct sum of subgroups of a given group or ideals of a given ring 239, 247

$D(f)$ the discriminant of the polynomial $f(x)$ 166

$\sigma_1, \sigma_2, \ldots, \sigma_n$ the elementary symmetric polynomials 218

$\sim$ the equivalence of matrices 40, 66

G/A a factor-algebra, factor-group, factor-ring 254, 235, 248

$P(N)$ the field obtained by adjoining the set N to the field P 249

$d(x) = (f(x), g(x))$ the greatest common divisor (g.c.d.) of the polynomials $f(x)$ and $g(x)$ 156

e the identity element of a group, ring, field 225, 244, 249

E the identity subgroup 229

$a \cap b$ the intersection of elements of a lattice 255

a^{-1} the inverse element for the element a of a group, ring, field 225, 245, 249

$\mathbf{A} \times \mathbf{B}$ the Kronecker product of matrices 101

$\tilde{D}_k$ the kth associated determinant 27

$\hat{D}_k$ the kth reciprocal (or adjoint) determinant 27

gA, $g+A$ the left coset with respect to the subgroup A, determined by the element g 231

$(a)_l$ the left ideal generated by the element a of a ring 246

$\mathbf{A}. \times \mathbf{B}$ the left Kronecker product of the matrices 102

$(a_{ij})_{mn}$ a matrix of dimensions $m \times n$ 52

$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}$, $(a_{ij})_1^n$ a matrix of the nth order xii, 52

M_{ij} the minor of the element a_{ij} 8

$A\begin{Bmatrix} i_1\, i_2 \dots i_k \\ j_1\, j_2 \dots j_k \end{Bmatrix}$ a minor of the matrix **A** 31

$S(ax_1^{k_1} x_2^{k_2} \dots x_n^{k_n})$ a monogenous polynomial with highest term $ax_1^{k_1} x_2^{k_2} \dots x_n^{k_n}$ 218

$W(a)$ the number of changes of sign in the series of values of the polynomials of the Sturm series, when $x = a$ 197

$\begin{pmatrix} 1\,2\,3\,4\,5\,6 \\ 6\,5\,1\,4\,2\,3 \end{pmatrix}$ a permutation 2

(163) (25) (4) a permutation written in cycles 3

$S_k = x_1^k + x_2^k + \dots + x_n^k$ a power sum 220

$\tilde{\mathbf{A}}_p$ the pth associated matrix 99

$A(\mathbf{x}; \mathbf{y})$ a quadratic form 117

$\hat{D}$ the reciprocal (or adjoint) determinant 25

$\hat{\mathbf{A}}$ the reciprocal (or adjoint) matrix 101

$R(f, g)$ the resultant of the polynomials $f(x)$ and $g(x)$ 165

Ag, $A+g$ the right coset with respect to the subgroup A, determined by the element g 231

$(a)_r$ the right ideal generated by the element a of a ring 246

$\mathbf{A} \times . \mathbf{B}$ the right Kronecker product of the matrices 102

$\mathbf{A}'$ the transposed matrix 6

(i, j), (i_1, i_2) transpositions 237

$g^{-1}Ag$ a subgroup conjugate to the subgroup A 234

$\{M\}$ the subgroup generated by the set of elements M 230–1

$\{A, B, \dots\}$ the subgroup generated by the subgroups A, B, ... 231

S_n the symmetric group of degree n 226

(a) the two-sided ideal generated by the element a of a ring 246

$a \cup b$ the union of elements of a lattice 255

0 the zero element of a group, ring 225, 244

SUBJECT INDEX

Adjunction of an element to a field 249
Algebra 251
  division 251
  Lie 251
  of finite dimension 251
  of quaternions 252
Algebraic closure xi
Algebraic complement of a minor 17
Algebraic element 250
Algebraic number 248
Algebraic operation 224
Algorithm
  Euclid's 157
  Routh's 207
Annihilator polynomial of a matrix 78
Arithmetical root of a matrix 113
Associative operation 224
Automorphism
  of a group 227
  of a partially ordered set 255

Bezout's theorem 154
  generalized 75
Bilinear form, *see* Form, bilinear
Binet–Cauchy theorem 22
Boolean algebra 259
Boolean ring 259
Budan–Fourier theorem 196

Cancellation in a ring 244
Canonical form of a bilinear form 126
Canonical form of a λ-matrix 67
Canonical λ-matrix 66
Canonical matrix 40
Canonical quadratic form 120
Cardan's formula 172
Cauchy's indices 204
Cauchy's theorem 232
Cayley's formulae 114
Cayley's table 226
Cayley's theorem 229
Cell, Jordan 88
Characteristic equation 76
  of a pencil of forms 141
Characteristic matrix 75
Characteristic minors of a system of linear equations 45
Characteristic numbers (roots) 75
  of an Hermitian matrix 110
  of an orthogonal matrix 104
  of a skew-Hermitian matrix 110
  of a skew-symmetric matrix 110
  of a symmetric matrix 110
  of a unitary matrix 104
Characteristic of a field 249
Characteristic polynomial 75–7
Characteristic values of a matrix 75
Characteristic vectors of a matrix 76
Circulant 23
  skew- 24
Cofactor 8
Common divisor of polynomials 156
Commutative operation 225
Commuting matrices 60–3
Complement in a lattice 259
Complete lattice 256

Component of an element in a direct factor 240
Conjugate elements of a group 234
Conjugate subgroups 234
Consistent system of linear equations 33
Continuant 15
Continuation of a direct decomposition 240
Contradictory system of linear equations 33
Coset (left, right) 231
Covering (in a lattice) 254
Cramer's rule 38
Criteria for the equivalence of λ-matrices 70
Criteria for the existence of a diagonal form of a matrix 84–5
Criteria of irreducibility of polynomials 182–5
Criterion
  Cohen's 183
  Eisenstein's 182
  Hurwitz's 214
  Kronecker–Capelli 45
  Liénard–Chipart 215
  Pólya's 183
  Routh's 214
  for similar matrices 81
  Sylvester's 131
Cubic resolvent 176
Cycle 3, 236
Cyclic group 229
Cyclic permutation 236
Cyclic subgroup 229

Decomposition
  of a group with respect to a subgroup 232
  of a polynomial into linear factors 180, 250
  of a polynomial into powers of a difference 155
Decrement of a permutation 3
Dedekind lattice 258
Degree
  of an element 229
  of a polynomial 152
    in n variables 216
  of a term of a polynomial in n variables 216
Derivative of a polynomial 155
Descartes' theorem 193
  stronger form 195
Determinant 4–6
  adjoint 25
    kth 27
  associated 26
    kth 27
  bordered 18
  echelon 18
  Hurwitz 210
  Jacobian 15
  of the order n 4
  of the order 2 4
  of the order 3 5
  of the product of matrices 55
  pseudosymmetric 28, 29
  reciprocal 25
    kth 27
  skew-symmetric 8, 28, 29
  symmetric 28
  Vandermonde 13
    generalized 14
Determinantal factor 68
Diagram 257
Difference of polynomials 152
Direct factor 240
Direct (Kronecker) product
  of determinants 29
  of matrices 101
Direct product of groups 239, 241
Direct sum
  of groups 239, 241
  of rings 247
Direct summand 240

Discriminant 166
of a quadratic form 118
Distributive property 243
Division
of matrix polynomials 74
with remainder 153
without remainder 153
Divisor of zero 244
Domain
integral 244
of imprimitivity 238
of stability 215
Duality 255

Echelon form of a system of linear equations 34
Eisenstein's criterion 182
Element
additive inverse 225
algebraic 250
of a ring, prime (irreducible) 247
of finite order 229
of infinite order 229
transcendental 250
Elementary divisors 71–3
of a cellular diagonal matrix 73
Elementary (fundamental) symmetric polynomials 218
Elementary λ-matrices 69
Elementary transformation
of a λ-matrix 66
of a matrix 40
of a system of linear equations 34
Endomorphism
of a group 228
Equality of polynomials 152, 216
Equation
binomial 178
biquadratic 178
of the fourth degree 176
of the second degree 171
of the third degree 172
Equivalent λ-matrices 66
Equivalent matrices 40
Equivalent quadratic forms 120
Equivalent systems of linear equations 33
Euclid's algorithm 157
Euler's method (of solving equations of the fourth degree) 177
Even combination 1
Even permutation 2
Extension
algebraic 250
of a field 249
simple 249
simple algebraic 250
simple transcendental 250

Factor of a polynomial, k-tuple 182
Factor-algebra 254
Factor-group 235
Factor-ring 248
Ferrari's method 176
Fibonacci series 16
Field 248
of algebraic numbers 248
algebraically closed 250
of characteristic zero 249
decomposition, of a polynomial 250
of residue classes 248
simple 249
Form
bilinear 117–8
Hermitian 148
polar 118
of the second kind 146
skew-symmetric 127
symmetric 118
(of a homogeneous polynomial) 216
of a matrix, natural normal 92
quadratic 118
Formula
Cardan's 172

Formula
Lagrange's (interpolation) 168
Newton's (interpolation) 169
Sylvester's 27
Formulae
Cayley's 114
Newton's, for power sums 220
Schur's 59
Vieta's 164
Waring's (first and second) 221
Four-group, Klein's 238
Free unknowns 35, 45
Frobenius' rule 129
Frobenius' theorem 253
on Hänkel forms 151
Fundamental (elementary) symmetric polynomials 218
Fundamental system of solutions 48
Fundamental theorem
on Abelian groups with a finite number of generators 242
of the algebra of complex numbers 162
on the expansion of a rational fraction into partial fractions 188
on symmetric polynomials 218

Gauss's method 34
General solution of a system of linear equations 35, 46
Generating elements of a group 231
Generation of a subgroup
by elements 230
subgroups 231
Greatest common divisor
determination 156, 157, 182
of several polynomials 161
of two polynomials 156
Group 224, 225–6
Abelian 225
indecomposable 242
additive 225
of a ring 244
alternating 229
commutative 225
cyclic 229
infinite 229
finite 226
Hamilton 235
multiplicative 225
of a field 249
of permutations 236
imprimitive 238
intransitive 238
primitive 238
transitive 238
of type p^∞ (quasi-cyclic) 231
quasi-cylic (of type p^∞) 231
simple 234
symmetric 226
with finite number of generators 230

Hadamard's inequality 30–31
generalized 31–2
Hamilton–Cayley theorem 77
Hamilton group 235
Hänkel form 149
Hänkel matrix 150
Hermitian bilinear form 148
Hermitian matrix 109
Hermitian quadratic form 148
Highest term of a polynomial 217
Hildenfinger's rule 129
Homogeneous polynomial (of a form) 216
Homomorphic mapping, natural 236
Homomorphism
of algebras 253
of fields 249
of groups 228
of rings 245
Horner's scheme 153

Hurwitz determinant 210
Hurwitz matrix 210
Hurwitz polynomial 214
Hurwitz's criterion 214
Hypercomplex system 251

Ideal
  generated by a given set 246
  of an algebra (left, right, two-sided) 253
  of a ring, principal 246
  of a ring (left, right, two-sided) 246
Identity
  Jacobi's 243
  of a field 249
  of a group 225
  of a lattice 256
  of a ring 244
Increasing sequence of subgroups 231
Indecomposable element of a ring 247
Independent cycles 3, 237
Index of a subgroup 232
Indices
  Cauchy's 204
  of inertia 128
Indirect (second) diagonal of a matrix xii
Inequality
  Hadamard's 30–1
    generalized 31–2
  Sylvester's 43
Interpolation formula
  Lagrange's 168
  Newton's 169
Intersection (in a lattice) 256
Invariant factors 67
Inverse element
  in a field 249
  in a group 225
  in a ring 245
  λ-matrices 70
  permutation 226
Inversion 1
Isomorphism
  of algebras 253
  of groups 227
  of lattices 256
  of partially ordered sets 255
  of rings 245

Jacobian determinants 15
Jacobian matrix 97
Jacobi's identity 243
Jacobi's method 122
Jacobi's theorem 129
Jordan cell 88
Jordan form of a matrix 89
Jordan matrix 88
  generalized 96

Kernel of a homomorphism 228, 246
Klein four-group 238
Kronecker–Capelli criterion 45
Kronecker product
  of determinants 29
  of matrices 101, 102

Lagrange's formula (interpolation) 168
Lagrange's method 121
Lagrange's theorem 232
Laplace's theorem 17
Lattice 255, 257
  complete 256
  Dedekind 258
  distributive 257
  modular 258
  with complements 259
Law
  associative 224
  commutative 225
  Cramer's 38

Law
distributive 243
Frobenius' 129
Hildenfinger's 129
of inertia for quadratic forms 128
Sarrius' 5
triangle 5
Left decomposition of a group with respect to a subgroup 232
Length of a cycle 237
Lexicographical expression for a polynomial 217
Liénard and Chipart criterion 215
Limits of the roots of a polynomial 190
Linear combination of rows 6
Linear form 130
Linear transformation of unknowns 53
Linearly ordered set 255
λ-matrix over a field 66

Maclaurin's method 191
Matrices
commuting 60–3
equivalent 40
similar 81
Matrix xi
adjoint 78, 101
associated 101
augmented 45
of a bilinear form 117, 146
canonical 40
cellular diagonal 73
characteristic 75
companion 92
conjugate 104
diagonal 72
echelon 40
Hänkel 150
Hermitian 109
Hurwitz 210
identity 56
inverse 56
for the product of matrices 56
involutory 79
Jacobian 97
normal 98
Jordan 88
generalized 95
lower triangular 64
nilpotent 77
non-singular 56
normal 115
canonical 116
orthogonal 103
canonical 116
polynomial 66
positive definite 31
pth associated 99
of a quadratic form 117
quasi-natural 94
reciprocal 101
reduced adjoint 80
scalar 60
singular 56
skew-Hermitian 109
skew-symmetric 109
square xi
symmetric 109
of a system of equations 45
triangular 64
unitary 104
upper triangular 64
with non-negative spectrum 133
with positive spectrum 113
zero 54
Method
Euler's, of solving an equation of the fourth degree 177
Ferrari's, of solving an equation of the fourth degree 176
for reduction to triangular form 10
Gauss's 34
Jacobi's 122
Lagrange's 121

Maclaurin's, of finding the limits of the real roots of a polynomial 190–1
Newton's, of finding the limits of the real roots of a polynomial 191
of quadratic forms 200
of recurrence relations 11
special cases 12
Sturm's 196
Methods of calculating the inverse matrix 57–8
Minimal polynomial of a matrix 78
Minor
complementary 16
of an element 8
of the kth order 16, 31
of the product of matrices 55
Modular (Dedekind) lattice 258
Modulus of a matrix 113
Multiplication
of determinants 21
of matrices 52–3
of permutations 226
Multiplicity of a root 162
Mutually prime polynomials 158

Negative definite quadratic form 130
Negative power of an element of a group 230
Newton's formula (interpolation) 169
Newton's formulae for power sums 220
Newton's method 191
Nodes of the u-line 98
Norm of a matrix 113
Normal divisor of a group 233
Normal form of a λ-matrix 67
Normal form of a quadratic form 126, 128
Normal Jacobian matrix 97
Normal λ-matrix 66
Normed (unit) vector 103
Number field xi

One–one correspondence 227
Order
of an element 229
of a finite group 226
of a matrix xi
Ordered set 254
Orthogonal matrix 103
canonical 105
geometrical meaning 108–9
Orthogonal transformation of variables 132
Orthogonal vectors 103
Orthogonally equivalent quadratic forms 139
Orthogonally similar matrices 104

Pair of forms 139
Partially ordered set 254
Pencil of quadratic forms 141
Permutation 2
cyclic 236
inverse 226
Pfaffian aggregate 28
Polar decomposition of a matrix 114
Polynomial 152
characteristic 75–6
Hurwitz 214
in n variables 216
irreducible 181
matrix 73
monogenous 218
of zero degree 152, 216
reducible 181
symmetric 217
Positive definite quadratic forms 130–2
Powers of a matrix 60

Power sums 220
expressed in terms of the elementary symmetric polynomials 220–1
Principal
(first) diagonal of a matrix xii
ideal 246
matrix of a pencil 141
minor of a matrix 43
of a system of linear equations 45
unknowns 35, 45
Product
of a matrix and a number 55
of matrices 52–3
of permutations 226
of polynomials 152
in n variables 217

Quadratic form 117
of definite sign 130
degenerate 118
Hermitian 148
non-degenerate 118
positive definite 130
of the second kind 146
semi-definite 131
singular 118
Quadratic forms, congruent 120
Quasi-natural form of a matrix 94
Quaternions 252
Quotient after division
of elements of a field 248
of polynomials 153
Quotient of matrices 61
solution 46

Rank
of a bilinear form 117, 146
of a matrix 40, 42–4
of a product of matrices 43
of a quadratic form 118
of a sum of matrices 42
Rational fraction 186
irreducible 187
proper 187
simple 190
Rational function 186
of a matrix 61
Rational roots of a polynomial (method of finding) 178–180
Reduced quadratic equation 171
Reduced system of equations 51
Reduction
to canonical form of a bilinear form 126
to canonical form of a polynomial of the second degree 135
to diagonal form of a quadratic form 120–1
to diagonal form of several quadratic forms (simultaneously) 135
of a Hermitian quadratic form to principal axes 149
of a quadratic form to principal axes 135
Regular case (in Routh's scheme) 207
Regular matrix polynomial 73
Regular pencil of quadratic forms 141
Remainder after division of polynomials 153
Resultant 165
Right decomposition of a group with respect to a subgroup 233
Ring 243
associative 243
Boolean 259
commutative 243
Lie 243
principal ideal 247
simple 246
with identity 244
without divisors of zero 244

Root
of a polynomial 162
of a matrix, arithmetical 113
of multiplicity k 162
simple 162
Routh–Hurwitz theorem 210
Routh's algorithm 207
Routh's criterion 214
Routh's scheme 207
Ruffini–Abel theorem 178

Sarrius' rule 5
Scalar product of vectors 103
Scheme
Horner's 153
Routh's 207
Schur's formula 59
Schur's theorem 109
Separation of multiple roots 163
Series
Fibonacci 16
Sturm 196
in a segment 196
Set
linearly ordered 255
ordered 255
partially ordered 254
Signature 128
Similar matrices 81
Skew field 251
of quaternions 251
Solution
of equations by radicals 178
of a system of linear equations 33
Space of coefficients 215
Spectrum
of the Kronecker product of matrices 103
of a matrix 76
of a polynomial in a matrix 77
Steinitz's theorem 250
Sturm's method 196
Sturm's series 196
Sturm's theorem 196
Sub-algebra 253
of a Boolean algebra 259
Subfield 249
Subgroup 229
cyclic 229
identity 229
invariant 233
generated by a given set of elements 230
generated by given subgroups 231
proper 229
Sub-lattice 256
Sub-ring 246
Sum
of matrices 52
of polynomials 152
in n variables 217
Sylvester's criterion 131
Sylvester's formula 27
Sylvester's inequality 43
Sylvester's theorem 27
Symmetric polynomial 217
expressed in terms of the elementary polynomials 218
System
of algebraic equations 221
of linear equations 33
determinate 33
inconsistent 33
indeterminate 33
of generators 230

Table, Cayley's 226
multiplication for an algebra of finite dimension 252
Theorem
Bezout's 154
generalized 75
Binet–Cauchy 22
Budan–Fourier 196

Theorem
Cauchy's 232
Cayley's 229
Descartes' 193
strengthened form 195
Frobenius' 253
on Hänkel forms 151
Hamilton–Cayley 77
Jacobi's, on the indices of inertia 129
Lagrange's 232
Laplace's 17
on Abelian groups with finite number of generators (fundamental) 242
symmetric polynomials (fundamental) 218
on the algebra of complex numbers (fundamental) 162
on the decomposition of rational fractions into partial fractions (fundamental) 188
on the homomorphism of groups 236
on the homomorphism of rings 248
on the rank of a matrix 41
Routh–Hurwitz 210
Ruffini–Abel 178
Schur's 109
Sturm's 196
Steinitz' 250
Sylvester's 27
Trace of a matrix 76
Transcendental element 250
Transformation by an element of a group 81, 234, 238
Transposed matrix 6
Transposition of the product of matrices 54
Transpositions 1, 237
Triangular system of linear equations 35

Unimodular λ-matrices 70
Union
of an increasing sequence of subgroups 231
of elements (in a lattice) 256
Unit of a ring 245
Unitarily similar matrices 104
Unitary matrix 104

Value of the polynomial of a Jordan matrix 91
of the polynomial of a matrix 61
Vandermonde determinant 13
Vieta's formulae 164

Waring's formulae 221

Zero
of a group 225
of a lattice 256
of a ring 244

OTHER TITLES IN THE SERIES IN PURE AND APPLIED MATHEMATICS

Vol. 1. WALLACE—*An Introduction to Algebraic Topology*
Vol. 2. PEDOE—*Circles*
Vol. 3. SPAIN—*Analytical Conics*
Vol. 4. MIKHLIN—*Integral Equations*
Vol. 5. EGGLESTON—*Problems in Euclidean Space: Application of Convexity*
Vol. 6. WALLACE—*Homology Theory on Algebraic Varieties*
Vol. 7. NOBLE—*Methods Based on the Wiener-Hopf Technique for the Solution of Partial Differential Equations*
Vol. 8. MIKUSIŃSKI—*Operational Calculus*
Vol. 9. HEINE—*Group Theory in Quantum Mechanics*
Vol. 10. BLAND—*The Theory of Linear Viscoelasticity*
Vol. 11. KURTH—*Axiomatics of Classical Statistical Mechanics*
Vol. 12. FUCHS—*Abelian Groups*
Vol. 13. KURATOWSKI—*Introduction to Set Theory and Topology*
Vol. 14. SPAIN—*Analytical Quadrics*
Vol. 15. HARTMAN and MIKUSIŃSKI—*The Theory of Lebesgue Measure and Integration*
Vol. 16. KULCZYCKI—*Non-Euclidean Geometry*
Vol. 17. KURATOWSKI—*Introduction to Calculus*
Vol. 18. GERONIMUS—*Polynomials Orthogonal on a Circle and Interval*
Vol. 19. ELSGOLC—*Calculus of Variations*
Vol. 20. ALEXITS—*Convergence Problems of Orthogonal Series*
Vol. 21. FUCHS and LEVIN—*Functions of a Complex Variable, Volume II*
Vol. 22. GOODSTEIN—*Fundamental Concepts of Mathematics*
Vol. 23. KEENE—*Abstract Sets and Finite Ordinals*
Vol. 24. DITKIN and PRUDNIKOV—*Operational Calculus in Two Variables and its Applications*
Vol. 25. VEKUA—*Generalized Analytic Functions*
Vol. 26. FASS and AMIR-MOÉZ—*Elements of Linear Spaces*
Vol. 27. GRADSHTEIN—*Direct and Converse Theorems*
Vol. 28. FUCHS—*Partially Ordered Algebraic Systems*
Vol. 29. POSTNIKOV—*Foundations of Galois Theory*
Vol. 30. BERMANT—*A Course of Mathematical Analysis, Part II*
Vol. 31. ŁUKASIEWICZ—*Elements of Mathematical Logic*
Vol. 32. VULIKH—*Introduction to Functional Analysis for Scientists and Technologists*
Vol. 33. PEDOE—*An Introduction to Projective Geometry*

Vol. 34. TIMAN—*Theory of Approximation of Functions of a Real Variable*
Vol. 35. CSÁSZÁR—*Foundations of General Topology*
Vol. 36. BRONSHTEIN and SEMENDYAYEV—*A Guide-Book to Mathematics for Technologists and Engineers*
Vol. 37. MOSTOWSKI and STARK—*Introduction to Higher Algebra*
Vol. 38. GODDARD—*Mathematical Techniques of Operational Research*
Vol. 39. TIKHONOV and SAMARSKII—*Equations of Mathematical Physics*
Vol. 40. MCLEOD—*Introduction to Fluid Dynamics*
Vol. 41. G. MOISIL—*The Algebraic Theory of Switching Circuits*
Vol. 42. OTTO—*Nomography*
Vol. 43. RANKIN—*An Introduction to Mathematical Analysis*
Vol. 44. BERMANT—*A Course of Mathematical Analysis, Part I*
Vol. 45. KRASNOSEL'SKII—*Topological Methods in the Theory of Nonlinear Integral Equations*
Vol. 46. KANTOROVICH and AKILOV—*Functional Analysis in Normed Spaces*
Vol. 47. JONES—*The Theory of Electromagnetism*
Vol. 48. FEJES TÓTH—*Regular Figures*
Vol. 49. YANO—*Differential Geometry on Complex and Almost Complex Spaces*
Vol. 50. MIKHLIN—*Variational Methods in Mathematical Physics*
Vol. 51. FUCHS and SHABAT—*Functions of a Complex Variable and Some of their Applications, Vol. I*
Vol. 52. BUDAK, SAMARSKII and TIKHONOV—*A Collection of Problems on Mathematical Physics*
Vol. 53. GILES—*Mathematical Foundations of Thermodynamics*
Vol. 54. SAUL'YEV—*Integration of Equations of Parabolic Type by the Methods of Nets*
Vol. 55. PONTRYAGIN *et al.*—*The Mathematical Theory of Optimal Processes*
Vol. 56. SOBOLEV—*Partial Differential Equations of Mathematical Physics*
Vol. 57. SMIRNOV—*A Course of Higher Mathematics, Vol. I*
Vol. 58. SMIRNOV—*A Course of Higher Mathematics, Vol. II*
Vol. 59. SMIRNOV—*A Course of Higher Mathematics, Vol. III, Part 1*
Vol. 60. SMIRNOV—*A Course of Higher Mathematics, Vol. III, Part 2*
Vol. 61. SMIRNOV—*A Course of Higher Mathematics, Vol. IV*
Vol. 62. SMIRNOV—*A Course of Higher Mathematics, Vol. V*
Vol. 63. NAIMARK—*Linear Representations of the Lorentz Group*
Vol. 64. BERMAN—*A Collection of Problems on a Course of Mathematical Analysis*
Vol. 65. MESHCHERSKII—*A Collection of Problems of Mechanics*
Vol. 66. ARSCOTT—*Periodic Differential Equations*
Vol. 67. SANSONE and CONTI—*Non-linear Differential Equations*
Vol. 68. VOLKOVYSKII, LUNTS and ARAMANOVICH—*A Collection of Problems on Complex Analysis*
Vol. 69. LYUSTERNIK and YANPOL'SKII—*Mathematical Analysis—Functions, Limits, Series, Continued Fractions*

Vol. 70. KUROSH—*Lectures in General Algebra*

Vol. 71. BASTON—*Some Properties of Polyhedra in Euclidean Space*

Vol. 72. FIKHTENGOL'TS—*The Fundamentals of Mathematical Analysis, Vol. 1*

Vol. 73. FIKHTENGOL'TS—*The Fundamentals of Mathematical Analysis, Vol. 2*

Vol. 74. PREISENDORFER—*Radiative Transfer on Discrete Spaces*

Vol. 75. FADDEYEV and SOMINSKII—*Elementary Algebra*

Vol. 76. LYUSTERNIK, CHERVONENKIS and YANPOLSKII—*Handbook for Computing Elementary Functions*

Vol. 77. SHILOV—*Mathematical Analysis—A Special Course*

Vol. 78. DITKIN and PRUDNIKOV—*Integral Transforms and Operational Calculus*

Vol. 79. POLOZHII—*The Method of Summary Representations for Numerical Solution of Problems of Mathematical Physics*